

Universidad de Oviedo

Departamento de Ingeniería Eléctrica, Electrónica, de Computadores y
Sistemas

Programa de Doctorado: Control de Procesos, Electrónica Industrial
e Ingeniería Eléctrica

TESIS DOCTORAL

Análisis y predicción de perfiles de consumo energético en edificios públicos mediante técnicas de minería de datos

Antonio Morán Álvarez

Septiembre 2012



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: ANÁLISIS Y PREDICCIÓN DE PERFILES DE CONSUMO ENERGÉTICO EN EDIFICIOS PÚBLICOS MEDIANTE TÉCNICAS DE MINERÍA DE DATOS.	Inglés: ANALYSIS AND FORECASTING OF ENERGY CONSUMPTION PROFILES IN PUBLIC BUILDINGS USING DATA MINING TECHNIQUES.
2.- Autor	
Nombre: ANTONIO MORÁN ÁLVAREZ	
Programa de Doctorado: CONTROL DE PROCESOS, ELECTRÓNICA INDUSTRIAL E INGENIERÍA ELÉCTRICA	
Órgano responsable: DEPARTAMENTO DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, DE COMPUTADORES Y SISTEMAS	

RESUMEN (en español)

El consumo sostenible y responsable de la energía requiere establecer medidas que persigan una mayor eficiencia energética. Un elevado conocimiento del consumo permite determinar la existencia de consumos innecesarios, optimizar la demanda diaria, balancear las cargas, etc. logrando reducir el nivel de consumo. Si, además, se tiene en cuenta la liberalización del mercado energético, el conocimiento extraído se puede aprovechar para mejorar la negociación con la compañía suministradora, reduciendo así costes. En esta tesis se ha realizado un análisis energético de una gran infraestructura, como es de la Universidad de León, caracterizada por tener varios puntos de medida (30 edificios) de elevado consumo y que están distribuidos, convirtiendo el análisis de datos en una tarea compleja.

A lo largo de este trabajo, se propone el desarrollo de una infraestructura tecnológica para la obtención de datos relacionados con el consumo energético, el almacenamiento y procesamiento de los mismos y su posterior visualización por parte de un usuario final. Para ello se implementa una arquitectura de triple capa, transparente y modular, aplicable al análisis de diferentes tipos de consumos energéticos como sistemas eléctricos, de calefacción o refrigeración. Además, se recogen datos de variables ambientales que influyen en el nivel de consumo y que serán útiles en las fases de análisis, interpretación de resultados y predicción de consumos.

Esta infraestructura incorpora herramientas de análisis con el fin de generar visualizaciones y modelos que permitan extraer el conocimiento implícito en los datos energéticos. Mediante el uso de la variable más representativa de consumo, que en el caso de esta tesis es la potencia activa, se generan perfiles con la dinámica diaria de cada edificio. El análisis de los perfiles o curvas de consumo por medio de técnicas de reducción de la dimensionalidad como son el CCA, SNE y t-SNE, permiten la obtención de representaciones en proyecciones 2D de los perfiles para realizar un análisis visual y una comparación de los datos. De esta forma se puede establecer qué edificios tienen un comportamiento similar y cuales se desvían de un comportamiento normal. El análisis de estas curvas y las variables ambientales determina cuales tienen una mayor influencia en el consumo. El uso de estas variables junto con otras indicadoras del nivel de consumo, como un índice de actividad que modela el nivel de uso de un edificio, permite crear modelos, mediante técnicas como el SOM, que representen los niveles de consumo y permitan una comparación entre los mismos.

Se han propuesto métodos para la predicción del consumo eléctrico a corto y medio plazo. Los modelos obtenidos hacen posible la estimación de los consumos de un edificio a medio plazo (1 mes) por medio de la relación de las variables ambientales con los perfiles de consumo. A partir de las estimaciones futuras de variables ambientales se puede predecir el consumo del edificio y establecer unos umbrales entre los que va a oscilar el consumo.

En el caso de la predicción a corto plazo se utilizan técnicas de reducción de la dimensionalidad, o variantes de las mismas, junto con técnicas de predicción propiamente dichas. Dado que existen diferentes posibilidades, en esta tesis, se plantean cuatro técnicas (CCA+ARIMA, SOM MIMO, KSOM y *Neural Gas*+SVR) y se comparan entre sí para ver cuál es la más adecuada. El uso de estas técnicas implica el uso de regresores (variables de entrada al modelo de predicción) que incorporan información de la evolución de los consumos. Para lograr una mayor precisión es necesario una búsqueda del regresor más adecuado con técnicas de selección, como el *k*-NN, que no requieren el entrenamiento de todo el algoritmo de predicción con el fin de ahorrar tiempo. En función de la dinámica del edificio y de cómo estos han sido agrupados, el regresor seleccionado cambiará, así como la precisión del método de predicción.

Para completar el estudio de los consumos, se analizan también los datos de las facturas de la compañía



suministradora, con el fin de comprobar el comportamiento de los puntos de facturación con respecto a los edificios. Este análisis da información de cómo la compañía penaliza a la Universidad por un consumo indebido o diferente al contratado y permite establecer de medidas de actuación, como la corrección del factor de potencia o de la potencia contratada.

Finalmente, con el propósito de que los métodos de análisis de datos y predicción propuestos puedan ser interpretados por una persona no versada en el análisis de datos, se han desarrollado una serie de herramientas de visualización, accesibles vía Internet, con el lenguaje *Processing*. Estas herramientas están dotadas de interactividad con el fin de modificar los datos a visualizar y así facilitar la extracción de conocimiento.

RESUMEN (en Inglés)

Sustainable and responsible consumption of energy requires establishing measures which pursue higher energy efficiency. A heavy knowledge of the consumption patterns enables to establish the existence of unnecessary consumption, optimize the daily demand, balance the loads, etc. succeeding in reducing the consumption level. If the energy market liberalization is also taken into account, the extracted knowledge can be leveraged to improve the negotiation with the supplier company, thereby reducing costs. This thesis analyze a large energy infrastructure, such as the University of León, characterized by having several measuring points (30 buildings) of high consumption which are scattered, so the data analysis becomes a complex task.

Throughout this paper, it is proposed the development of a technological infrastructure to gather data related to energy consumption, store and process them and show them later to an end user. This architecture implements a three-tier structure, transparent and modular, which can be used with different types of energy consumption such as electrical, heating or cooling systems. In addition, data are collected from environmental variables which have an influence on the level of consumption. They will be useful in the phases of analysis, interpretation of the results and forecasting.

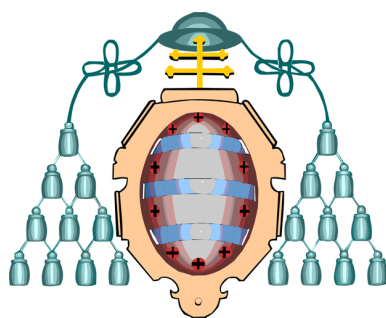
This infrastructure includes analysis tools to generate visualizations and models in order to extract implicit knowledge in the energy data. Using the most representative consumption variable, which is the active power in this thesis, profiles of the daily dynamic of each building are generated. The analysis of the profiles or curves of consumption by dimensionality reduction techniques such as the CCA, SNE and t-SNE allow 2D projections of the profiles to be obtained for visual analysis and data comparison. This way, it can be established which buildings have a similar behavior and which ones have deviations from normal behavior. The analysis of these curves and the environmental variables also determines which ones have a greater influence on consumption. It is possible to create models using these variables along with other indicators of the consumption level, such as an activity index that models the usage level of a building. These models can be created using techniques such as SOM, they represent the consumption levels and allow a comparison between them.

In addition, methods for short and medium term forecasting of electricity consumption have been proposed. The obtained models make it possible to estimate the consumption of a building in the medium term (1 month) using the relationship between environmental variables and consumption profiles. From the future estimation of environmental variables, it is possible to forecast the consumption of a building and establish thresholds for the consumption levels.

In the case of short-term prediction techniques, dimensionality reduction techniques, or modifications thereof, are used together with prediction techniques. Since there are different possibilities, in this thesis, four techniques are tested (CCA + ARIMA, SOM MIMO, KSOM and Neural Gas + SVR) and compared with each other to determine which is best. The use of these techniques involves the use of regressors (inputs to the prediction model) that incorporate information on the evolution of consumption. To achieve greater accuracy is required to choose the best regressor using selection techniques, such as k-NN, which does not need to train the algorithm, in order to save time. Depending on the dynamics of the building and how these have been grouped, the regressor is modified, and the accuracy of the prediction method changes.

To complete the consumption study, the data from the supplier company bills are also analyzed, in order to check the behavior of the billing points regarding to buildings. This analysis provides information about the company penalizations to the University for high or different consumption to the contracted one. Furthermore, it allows to establish performance measures, such as power factor correction or contracted power.

Finally, a set of visualization tools has been developed with Processing, and they are accessible via the Internet, so that the methods of data analysis and forecasting which have been proposed can be understood by an unskilled person in data analysis. These tools are interactive so the data to visualize can be modified and knowledge extraction becomes an easier task.



Universidad de Oviedo

Departamento de Ingeniería Eléctrica, Electrónica, de Computadores y
Sistemas

Programa de Doctorado: Control de Procesos, Electrónica Industrial
e Ingeniería Eléctrica

TESIS DOCTORAL

Análisis y predicción de perfiles de consumo energético en edificios públicos mediante técnicas de minería de datos

Memoria presentada para la obtención del grado de Doctor por la
Universidad de Oviedo

Autor: D. Antonio Morán Álvarez
Directores: Dr. D. Abel Alberto Cuadrado Vega
Dr. D. Juan José Fuertes Martínez

Gijón, Septiembre de 2012

A mis padres

“Sir, algún día podrá usted gravarla con impuestos.”

Michael Faraday

“El científico no tiene por objeto un resultado inmediato. Él no espera que sus ideas avanzadas sean fácilmente aceptadas. Su deber es sentar las bases para aquellos que están por venir, y señalar el camino.”

Nikola Tesla

Agradecimientos

Esta tesis ha visto la luz gracias a la ayuda y apoyo de un gran número de personas sin las cuales todo esto no habría sido posible. Podría escribir otra tesis sólo para agradecerse. En particular se lo quiero agradecer:

A Juan José Fuertes, director de esta tesis, por el apoyo y tutela que me ha venido ofreciendo no sólo a lo largo de la tesis, sino también desde el proyecto final de carrera. Gracias a él me inicié en el mundo de la minería de datos y de los procesos industriales.

A Manuel Domínguez, director del grupo SUPPRESS en el que he desarrollado la tesis, por abrirme la puertas al mundo de la investigación y brindarme la oportunidad de empezar a trabajar.

A mis compañeros de laboratorio, Serafin, Miguel Ángel, Pablo, Perfecto, Carlos y Sergio; por sus consejos, sugerencias y servicios prestados. No sólo han sido compañeros de trabajo, sino también amigos fuera de él.

Al grupo de investigación Supervisión y Diagnóstico de Procesos Industriales de la Universidad de Oviedo. En especial a Abel Cuadrado, por codirigir esta tesis y aportarme nuevos enfoques que permitieron perfeccionar este trabajo, y a Ignacio Díaz, por iniciarme en el mundo de la visualización y por sus ideas capaces de generar nuevos campos dentro de esta tesis.

Al grupo de investigación *Parsimonious Modelling* de *Aalto University School of Science* dirigido por Jaakko Hollmén, por acogerme durante mi estancia en su grupo y aportar nuevas ideas con las que mejorar esta tesis.

A la Junta de Castilla y León, que concedió la beca de investigación bajo la cual se desarrolló este trabajo.

A mis padres, por su apoyo incondicional, en especial en aquellos momentos en los que la tesis amenazaba con “ahogarme”.

A mis amigos, en especial a Javi y a Marta por su ayuda con la maquetación y correcciones, y a los compañeros de Astronomía por permitirme olvidar durante breves momentos el arduo trabajo en el que estaba inmerso.

Resumen

El consumo sostenible y responsable de la energía requiere establecer medidas que persigan una mayor eficiencia energética. Un elevado conocimiento del consumo permite determinar la existencia de consumos innecesarios, optimizar la demanda diaria, balancear las cargas, etc. logrando reducir el nivel de consumo. Si, además, se tiene en cuenta la liberalización del mercado energético, el conocimiento extraído se puede aprovechar para mejorar la negociación con la compañía suministradora, reduciendo así costes. En esta tesis se ha realizado un análisis energético de una gran infraestructura, como es de la Universidad de León, caracterizada por tener varios puntos de medida (30 edificios) de elevado consumo y que están distribuidos, convirtiendo el análisis de datos en una tarea compleja.

A lo largo de este trabajo, se propone el desarrollo de una infraestructura tecnológica para la obtención de datos relacionados con el consumo energético, el almacenamiento y procesamiento de los mismos y su posterior visualización por parte de un usuario final. Para ello se implementa una arquitectura de triple capa, transparente y modular, aplicable al análisis de diferentes tipos de consumos energéticos como sistemas eléctricos, de calefacción o refrigeración. Además, se recogen datos de variables ambientales que influyen en el nivel de consumo y que serán útiles en las fases de análisis, interpretación de resultados y predicción de consumos.

Esta infraestructura incorpora herramientas de análisis con el fin de generar visualizaciones y modelos que permitan extraer el conocimiento implícito en los datos energéticos. Mediante el uso de la variable más representativa de consumo, que en el caso de esta tesis es la potencia activa, se generan perfiles con la dinámica diaria de cada edificio. El análisis de los perfiles o curvas de consumo por medio de técnicas de reducción de la dimensionalidad como son el CCA, SNE y t-SNE, permiten la obtención de representaciones en proyecciones 2D de los perfiles para realizar un análisis visual y una comparación de los datos. De esta forma se puede establecer qué edificios tienen un comportamiento similar y cuales se desvían de un comportamiento normal. El análisis de estas curvas y las variables ambientales determina cuales tienen una mayor influencia en el consumo. El uso de estas variables junto con otras indicadoras del nivel de consumo, como un índice de actividad que modela el nivel de uso de un edificio, permite crear modelos, mediante técnicas como el SOM, que representen los niveles de consumo y permiten una comparación entre los mismos.

Adicionalmente, se han propuesto métodos para la predicción del consumo eléctrico a corto y medio plazo. Los modelos obtenidos hacen posible la estimación de los consumos de un edificio a medio plazo (1 mes) por medio de la relación de las variables ambientales

con los perfiles de consumo. A partir de las estimaciones futuras de variables ambientales se puede predecir el consumo del edificio y establecer unos umbrales entre los que va a oscilar el consumo.

En el caso de la predicción a corto plazo se utilizan técnicas de reducción de la dimensionalidad, o variantes de las mismas, junto con técnicas de predicción propiamente dichas. Dado que existen diferentes posibilidades, en esta tesis, se plantean cuatro técnicas (CCA+ARIMA, SOM MIMO, KSOM y *Neural Gas*+SVR) y se comparan entre sí para ver cuál es la más adecuada. El uso de estas técnicas implica el uso de regresores (variables de entrada al modelo de predicción) que incorporan información de la evolución de los consumos. Para lograr una mayor precisión es necesario una búsqueda del regresor más adecuado con técnicas de selección, como el k -NN, que no requieren el entrenamiento de todo el algoritmo de predicción con el fin de ahorrar tiempo. En función de la dinámica del edificio y de cómo estos han sido agrupados, el regresor seleccionado cambiará, así como la precisión del método de predicción.

Para completar el estudio de los consumos, se analizan también los datos de las facturas de la compañía suministradora, con el fin de comprobar el comportamiento de los puntos de facturación con respecto a los edificios. Este análisis da información de cómo la compañía penaliza a la Universidad por un consumo indebido o diferente al contratado y permite establecer de medidas de actuación, como la corrección del factor de potencia o de la potencia contratada.

Finalmente, con el propósito de que los métodos de análisis de datos y predicción propuestos puedan ser interpretados por una persona no versada en el análisis de datos, se han desarrollado una serie de herramientas de visualización, accesibles vía Internet, con el lenguaje *Processing*. Estas herramientas están dotadas de interactividad con el fin de modificar los datos a visualizar y así facilitar la extracción de conocimiento.

Abstract

Sustainable and responsible consumption of energy requires establishing measures which pursue higher energy efficiency. A heavy knowledge of the consumption patterns enables to establish the existence of unnecessary consumption, optimize the daily demand, balance the loads, etc. succeeding in reducing the consumption level. If the energy market liberalization is also taken into account, the extracted knowledge can be leveraged to improve the negotiation with the supplier company, thereby reducing costs. This thesis analyze a large energy infrastructure, such as the University of León, characterized by having several measuring points (30 buildings) of high consumption which are scattered, so the data analysis becomes a complex task.

Throughout this paper, it is proposed the development of a technological infrastructure to gather data related to energy consumption, store and process them and show them later to an end user. This architecture implements a three-tier structure, transparent and modular, which can be used with different types of energy consumption such as electrical, heating or cooling systems. In addition, data are collected from environmental variables which have an influence on the level of consumption. They will be useful in the phases of analysis, interpretation of the results and forecasting.

This infrastructure includes analysis tools to generate visualizations and models in order to extract implicit knowledge in the energy data. Using the most representative consumption variable, which is the active power in this thesis, profiles of the daily dynamic of each building are generated. The analysis of the profiles or curves of consumption by dimensionality reduction techniques such as the CCA, and t-SNE SNE allow 2D projections of the profiles to be obtained for visual analysis and data comparison. This way, it can be established which buildings have a similar behavior and which ones have deviations from normal behavior. The analysis of these curves and the environmental variables also determines which ones have a greater influence on consumption. It is possible to create models using these variables along with other indicators of the consumption level, such as an activity index that models the usage level of a building. These models can be created using techniques such as SOM, they represent the consumption levels and allow a comparison between them.

In addition, methods for short and medium term forecasting of electricity consumption have been proposed. The obtained models make it possible to estimate the consumption of a building in the medium term (1 month) using the relationship between environmental variables and consumption profiles. From the future estimation of environmental variables,

it is possible to forecast the consumption of a building and establish thresholds for the consumption levels

In the case of short-term prediction techniques, dimensionality reduction techniques, or modifications thereof, are used together with prediction techniques. Since there are different possibilities, in this thesis, four techniques are tested (CCA + ARIMA, SOM MIMO, KSOM and Neural Gas + SVR) and compared with each other to determine which is best. The use of these techniques involves the use of regressors (inputs to the prediction model) that incorporate information on the evolution of consumption. To achieve greater accuracy is required to choose the best regressor using selection techniques, such as k -NN, which does not need to train the algorithm, in order to save time. Depending on the dynamics of the building and how these have been grouped, the regressor is modified, and the accuracy of the prediction method changes.

To complete the consumption study, the data from the supplier company bills are also analyzed, in order to check the behavior of the billing points regarding to buildings. This analysis provides information about the company penalizations to the University for high or different consumption to the contracted one. Furthermore, it allows to establish performance measures, such as power factor correction or contracted power.

Finally, a set of visualization tools has been developed with *Processing*, and they are accessible via the Internet, so that the methods of data analysis and forecasting which have been proposed can be understood by an unskilled person in data analysis. These tools are interactive so the data to visualize can be modified and knowledge extraction becomes an easier task.

Índice general

1. Introducción	1
1.1. El problema de la energía	1
1.1.1. Consumo de energía	1
1.1.2. Minería de datos	3
1.1.3. Predicción	6
1.2. Objetivos de la tesis	7
1.3. Estructura de la tesis	8
2. Reducción de la dimensionalidad	9
2.1. Introducción a la reducción de la dimensionalidad	9
2.2. Reducción de la dimensionalidad por preservación de la distancia	12
2.2.1. Métodos de escalado multidimensional	14
2.2.2. Proyección de Sammon	15
2.2.3. Análisis de componentes curvilíneas	17
2.3. Reducción por preservación de la topología	19
2.3.1. <i>Locally Linear Embedding</i>	19
2.3.2. <i>Laplacian Eigenmap</i>	21
2.3.3. <i>Stochastic Neighbor Embedding</i>	22
2.3.4. Mapas autoorganizados	24
2.4. Cuantización y <i>clustering</i> de vectores	32
2.4.1. <i>k-means</i>	34
2.4.2. <i>Neural Gas</i>	35
2.4.3. Agrupación jerárquica	35
3. Predicción de series temporales	37
3.1. Series temporales	37
3.2. Estrategias de predicción de series temporales	38
3.2.1. Estrategias de salida única	38

	Índice general	IX
3.2.2. Estrategias de salida múltiple	40	
3.3. Técnicas de predicción	41	
3.4. Métodos basados en <i>clustering</i>	44	
3.5. Uso del SOM para predicción	45	
3.5.1. Modelo VQTAM	46	
3.5.2. SOM Recurrente	47	
3.5.3. SOM Recursivo	47	
3.5.4. <i>Double Vector Quantization Method</i>	48	
3.5.5. SOM Autorregresivos	49	
3.5.6. KSOM	50	
3.6. Predicción con técnicas de <i>manifold learning</i>	51	
4. Estructura del sistema de medida de consumos energéticos	54	
4.1. Edificios y puntos de medida	54	
4.2. Arquitectura de medida y supervisión	57	
4.3. Capa física	60	
4.3.1. Equipos físicos	63	
4.3.2. Arquitectura de red	66	
4.4. Capa intermedia	68	
4.4.1. Hardware	68	
4.4.2. Software	69	
4.5. Capa cliente	75	
5. Metodología	80	
5.1. Definición del espacio de entrada	80	
5.2. Proyección de los datos	83	
5.2.1. Selección de la técnica	85	
5.2.2. Mapas continuos	86	
5.2.3. Comparación de perfiles	89	
5.3. Modelado de los datos	89	
5.3.1. Generación de modelos	90	
5.3.2. Comparación de modelos de perfiles	91	
5.4. Predicción a largo plazo	93	
5.5. Predicción a corto plazo	97	
5.5.1. Predicción MIMO	97	
5.5.2. Predicción Directa	100	
6. Experimentación y resultados	110	

6.1. Obtención del espacio de entrada	111
6.1.1. Selección de la variable de consumo	111
6.1.2. Selección de las variables ambientales	112
6.2. Proyección de los perfiles de consumo	117
6.2.1. Selección de la proyección	117
6.2.2. Proyección de los edificios	124
6.2.3. Comparación mediante técnicas de reducción de la dimensionalidad . .	124
6.3. Modelado de los edificios	132
6.3.1. Generación de modelos	133
6.3.2. Comparación de modelos de perfiles	134
6.4. Predicción a largo plazo	136
6.5. Predicción a corto plazo	142
6.5.1. Predicción MIMO	143
6.5.2. Predicción Directa	144
6.5.3. Resultados de la predicción a corto plazo	148
6.6. Visualización de facturas	153
7. Conclusiones y líneas futuras	161
7.1. Conclusiones	161
7.2. Aportaciones	163
7.3. Líneas futuras	165
Bibliografía	168

Índice de tablas

4.1. Descripción de los edificios del campus de la Universidad de León.	56
4.2. Variables eléctricas capturadas por los equipos para cada edificio.	64
6.1. Correlaciones para los edificios analizados para temperatura (T), humedad (Hum), radiación solar (RS) y mínimo de la correlación horaria (min Corr h).	114
6.2. Variables para el cálculo del índice de actividad.	116
6.3. Variables del espacio de entrada.	118
6.4. <i>Clustering</i> de los edificios en base a la forma de la proyección de su dinámica.	132
6.5. Índices de predicción para todos los edificios de la universidad usando los modelos generados con el SOM.	140
6.6. Variables ambientales utilizadas para la predicción.	143
6.7. Índices de predicción para la medida del error de predicción a corto plazo.	153
6.8. Variables de las facturas empleadas para la reducción de la dimensión.	154
6.9. Ahorro anual por la corrección de la potencia contratada en las facturas de la Universidad.	160

Índice de figuras

2.1. Reducción de la dimensionalidad.	10
2.2. <i>Manifold Swiss-Roll</i>	11
2.3. Ejemplo de proyección PCA.	12
2.4. Comparación de los distintos tipos de distancia.	13
2.5. Proyección del <i>Swiss-Roll</i> con MDS e Isomap.	15
2.6. Proyección del <i>Swiss-Roll</i> con NLM y GNLM.	16
2.7. Proyección del <i>Swiss-Roll</i> con CCA y CDA.	18
2.8. Proyección del <i>Swiss-Roll</i> con LLE y LE.	20
2.9. Proyección del <i>Swiss-Roll</i> con SNE y t-SNE.	23
2.10. Proyección del mapa autoorganizado (SOM).	25
2.11. Esquema del proceso de adaptación en el mapa autoorganizado (SOM).	27
2.12. Matriz de distancias 2D y 3D.	30
2.13. Mapa de trayectorias y de transiciones.	31
2.14. Método de agrupamiento por <i>k-means</i> y jerárquico.	34
4.1. Esquema del suministro de electricidad para todos los edificios.	58
4.2. Arquitectura del sistema de medida y supervisión.	61
4.3. Diagrama de la arquitectura del sistema de medida y supervisión.	62
4.4. Configuración de la máquina virtual.	70
4.5. Supervisión tradicional de las facturas.	77
4.6. Supervisión avanzada de las facturas.	78
5.1. Codificación del índice de actividad.	81
5.2. Codificación de las variables temporales.	82
5.3. Obtención de las proyecciones de los perfiles.	84
5.4. Generación de mapas continuos	87
5.5. Obtención de los modelos utilizando el SOM.	92
5.6. Proceso de predicción a largo plazo.	95
5.7. Proceso de predicción MIMO con SOM.	98

5.8. Proceso de predicción utilizando CCA.	101
5.9. Proceso de predicción directa utilizando KSOM.	106
5.10. Proceso de predicción directa utilizando <i>Neural Gas</i> y SVR	108
6.1. Análisis de la dependencia de la potencia con la hora en el edificio de Filosofía.112	
6.2. Análisis de la dependencia de la potencia con la hora en el edificio de Animalario.113	
6.3. Correlaciones entre las variables meteorológicas y la potencia media diaria consumida para el edificio de Filosofía.	115
6.4. Correlaciones entre las variables meteorológicas y la potencia media diaria consumida para el edificio de Animalario.	115
6.5. Índice Davies-Bouldin para los distintos métodos de <i>clustering</i>	117
6.6. Comparación de los métodos de reducción de la dimensionalidad para los perfiles eléctricos diarios (I).	119
6.7. Comparación de los métodos de reducción de la dimensionalidad para los perfiles eléctricos diarios (II).	120
6.8. Índices de fiabilidad y continuidad para los métodos de reducción de la dimensionalidad.	122
6.9. Proyección SNE de los perfiles de consumo para el edificio de Filosofía.	125
6.10. Proyección SNE de los perfiles de consumo para el edificio de Animalario.	126
6.11. Comparación de perfiles de varios edificios usando SNE.	128
6.12. Comparación de perfiles de varios edificios usando t-SNE.	129
6.13. Calendario y gráfica de perfiles en <i>Processing</i>	129
6.14. Comparación del consumo de edificios en función del día de la semana.	130
6.15. Comparación del consumo de edificios en función del mes.	130
6.16. Comparación del consumo de los edificios en función de la elipse que les contiene.131	
6.17. Mapas de diferencias para comparación de los modelos de los edificios de Filosofía y Derecho.	134
6.18. Mapas de diferencias para comparación de los modelos de los edificios de Filosofía y Animalario.	135
6.19. Mapas de módulos de diferencias.	136
6.20. Predicción de consumos a largo plazo para el edificio de Filosofía.	138
6.21. Predicción de consumos a largo plazo para el edificio de Animalario.	138
6.22. Comparación del NMRSE en función de la variación de la temperatura con respecto a la real.	139
6.23. Herramienta de <i>Processing</i> para la visualización de la predicción de consumos.141	
6.24. Herramienta de <i>Processing</i> para la visualización de los modelos del SOM.	141
6.25. Selección del regresor para predicción directa correspondiente al edificio de Animalario.	146

6.26. Selección del regresor para predicción directa correspondiente al edificio de Filosofía.	146
6.27. Selección del regresor para predicción directa correspondiente al edificio de Veterinaria.	147
6.28. Resultados de predicción a corto plazo en Filosofía para el instante $t + 24$. . .	150
6.29. NRMSE para los distintos métodos en Filosofía.	150
6.30. Resultados de predicción a corto plazo en Veterinaria para el instante $t + 24$. . .	151
6.31. NRMSE para los distintos métodos en Veterinaria.	151
6.32. Resultados de predicción a corto plazo en Animalario para el instante $t + 24$. . .	152
6.33. NRMSE para los distintos métodos en Animalario.	152
6.34. Herramienta en <i>Processing</i> para la visualización de las facturas.	155
6.35. Visualización de la energía activa consumida en cada factura.	156
6.36. Visualización de la energía reactiva consumida en cada factura.	157
6.37. Visualización del $\cos \phi$ en cada factura.	159
6.38. Visualización de la potencia pico marcada por el maxímetro.	159

Lista de símbolos

\mathbb{R}^p	Espacio de entrada de p dimensiones.
\mathbb{R}^d	Espacio de salida de d dimensiones.
\mathbf{x}_i	Vector de entrada.
\mathbf{y}_i	Coordenadas en el espacio de salida.
$d(\mathbf{x}_i, \mathbf{x}_j)$	Distancia entre dos puntos de datos \mathbf{x}_i y \mathbf{x}_j .
$\ \cdot\ $	Distancia euclídea.
\mathbf{w}_i	Vector de pesos o <i>codebook</i> del SOM.
\mathbf{q}_i	Vector <i>codebook</i> de un <i>cluster</i> .
Q_k	<i>Cluster</i> o agrupamiento.
c	Neurona ganadora.
h_{ci}	Función de vecindad.
$\alpha(t)$	Velocidad o tasa de aprendizaje en el instante de tiempo t .
$\sigma(t)$	Radio de vecindad en el instante de tiempo t .
$\hat{x}(t + s)$	Muestra predicha en el instante $t + s$.
$\mathbf{x}(t + s)$	Regresor de predicción asociado al instante $t + s$.
a_h	Variable de consumo para la hora h .
\mathbf{a}_i	Vector de la curva diaria de consumo.
\mathbf{e}_i	Vector de variables ambientales.
Ω	Máscara de entrenamiento del SOM.
$\Phi(\mathbf{x})$	Núcleo de base radial. Función RBF.
σ	Factor de ancho en funciones gaussianas.

$r(i, j)$	Rango de la muestra j con respecto a la distancia i .
$S(\mathbf{x} \mathbf{x}_i, \mathbf{y}_i)$	Estimador de Nadaraya-Watson de la interpolación del punto \mathbf{x} dados los puntos \mathbf{x}_i e \mathbf{y}_i .
$p_i, p(\mathbf{x}_i)$	Función de densidad de probabilidad del vector \mathbf{x}_i .
$p_{i j}, p(\mathbf{x}_i \mathbf{x}_j)$	Densidad de probabilidad condicional entre el vector \mathbf{x}_i y \mathbf{x}_j .
N	Número de datos.
n	Dimensión de un vector de datos.
m	Dimensión de un regresor.
D, M, M_1, M_2	Tamaño de un regresor en días, horas.
H_x	Coordenada x (seno) de la variable temporal H .
H_y	Coordenada y (coseno) de la variable temporal H .
$f(\cdot)$	Función de predicción.
A_t	Valor actual de una serie.
F_t	Valor predicho de una serie.
E	Función de coste.

Lista de acrónimos

AR	<i>Autorregresivo</i>
ARMA	<i>Autoregressive Moving Average</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BMU	<i>Best Matching Unit</i>
CCA	<i>Curvilinear Component Analysis</i>
CDA	<i>Curvilinear Distance Analysis</i>
DBMS	<i>Database Management System</i>
DMZ	<i>De-Militarized Zone</i>
GNLM	<i>Geodesic Nonlinear Mapping</i>
GRNN	<i>General Regression Neural Network</i>
HTML	<i>Hypertext Markup Language</i>
HVAC	<i>Heating, Ventilating and Air Conditioning Systems</i>
Isomap	<i>Isometric Feature Mapping</i>
KR-SOM	<i>Kernel Regression SOM</i>
LAN	<i>Local Area Network</i>
LLM	<i>Local Linear Map</i>
LBG	<i>Linde-Buzo-Gray</i>
LE	<i>Laplacian Eigenmap</i>
LLE	<i>Locally Linear Embedding</i>
MDS	<i>Multidimensional Scaling</i>
MIMO	<i>Multiple Input - Multiple Output</i>
MLP	<i>Multilayer Perceptron</i>

MAPE	<i>Mean Absolute Percentage Error</i>
MSE	<i>Mean Square Error</i>
NLM	<i>Nonlinear Mapping</i>
NRMSE	<i>Normalized Root Mean Square Error</i>
OSI	<i>Open System Interconnection</i>
PCA	<i>Principal Component Analysis</i>
RBF	<i>Radial Basis Function Network</i>
ReIRMSE	<i>Relative Root Mean Square Error</i>
RMSE	<i>Root Mean Square Error</i>
SOM	<i>Self-Organizing Map</i>
SNE	<i>Stochastic Neighbor Embedding</i>
SVM	<i>Support Vector Machines</i>
SVR	<i>Support Vector Regression</i>
t-SNE	<i>t-Distributed Stochastic Neighbor Embedding</i>
VLAN	<i>Virtual Local Area Network</i>
VQ	<i>Vector Quantization</i>
VQTAM	<i>Vector-Quantized Temporal Associative Memory</i>

Introducción

En este capítulo introductorio se motiva el trabajo que se lleva a cabo a lo largo de esta tesis, se plantean los objetivos que se pretenden conseguir y se presenta la estructura de capítulos en los que se divide este documento.

1.1. El problema de la energía

1.1.1. Consumo de energía

El aumento del consumo de energía, junto con el aumento del coste de su producción, ha llevado en estos últimos años al desarrollo de estrategias para la gestión eficiente del consumo energético. En este sentido, los gobiernos desarrollan políticas que fomentan el ahorro y la eficiencia energética y penalizan el derroche de energía. En España, mediante el “Plan de Ahorro y Eficiencia Energética 2008-2012”, promovido por el Instituto para la Diversificación y Ahorro de la Energía (IDAE), se plantean una serie de medidas de ahorro, como son la instalación de sistemas de supervisión, modernización del equipamiento, control de la iluminación, etc. El objetivo final de estas directrices es lograr un sistema energético sostenible que concuerde con el desarrollo económico del país (Instituto para la Diversificación y Ahorro de la Energía, 2007).

Uno de los sectores menos eficientes en el consumo energético es el sector terciario y de servicios, debido a la dificultad de no poder controlar sus consumos. Se estima que el 30 % de la energía consumida en España se debe a este sector, a esto hay que añadir que la demanda energética requerida por este sector ha sufrido un gran crecimiento a lo largo de los últimos años (US Energy Information Administration, 2007). Este aumento se debe fundamentalmente al incremento en la cantidad de equipos instalados en los edificios destinados a aumentar los servicios y el nivel de comfort. El consumo de energía en el sector terciario se debe, sobre todo, a los sistemas de iluminación, refrigeración, calefacción, ventilación y aire acondicionado (*Heating, Ventilating and Air Conditioning Systems*, HVAC) (Bertoldi y Atanasiu, 2009). Dado que este equipamiento no suele pertenecer a la gama alta de eficiencia energética, es necesario adoptar directrices e instaurar acciones concretas dirigidas a aprovechar el gran potencial de ahorro de energía en los edificios que aún está sin explotar (Unión Europea, 2010).

Debido a su gran influencia en el nivel de consumo, los edificios del sector terciario son objeto de numerosas políticas, estrategias, planes o acciones seguidas con el fin de mejorar su eficiencia energética. El ahorro de energía en edificios puede lograrse mediante una mejora de su gestión energética, equipos e instalaciones, como por ejemplo supervisar y analizar los consumos energéticos y sus desviaciones, definir un protocolo de puesta en marcha y parada de los sistemas de climatización e iluminación, incorporar equipos de bajo consumo y alto rendimiento, etc.

El ahorro en el gasto en energía no solo puede venir por la reducción del consumo, sino también por la reducción del importe de las facturas. En el caso de las facturas eléctricas existe toda una serie de penalizaciones debidas a desvíos en el consumo que pueden ser corregidos, bien sea modificando los patrones de consumo, o bien negociando las condiciones del contrato de suministro.

En España, la ley 54/1997 del sector eléctrico (España, 1997) estableció un proceso gradual para la entrada de consumidores en el mercado libre de la electricidad. Esta ley permite acudir a un mercado liberalizado en el cual el cliente contrata la electricidad con una empresa comercializadora a un precio y condiciones libremente pactadas, acude al mercado mayorista o *pool* de energía, o bien suscribe un contrato directamente con un productor. Las ventajas son interesantes desde el punto de vista económico, ya que se puede negociar el precio de la energía eléctrica con una empresa comercializadora, o bien solicitar ofertas a varias compañías.

Los niveles de consumo de energía de los edificios públicos y sobre todo de los destinados a la docencia e investigación son elevados y en muchos casos poseen suministros de energía eléctrica en media tensión. Por este motivo, las administraciones o instituciones públicas deberían acudir al mercado liberalizado para la contratación de electricidad, lo que implica que previamente se debe adquirir un perfecto conocimiento del comportamiento eléctrico de los edificios para, posteriormente, proceder de forma favorable en la negociación.

Para poder obtener este conocimiento del comportamiento, no solo de la evolución del consumo eléctrico, sino también de cualquier otro tipo de variable energética, se requiere un sistema de supervisión básico que permita medir, registrar, visualizar y analizar estas variables (Rossiter, 2005). En muchos casos, estas tareas son automatizadas por medio de sistemas SCADA (*Supervisory Control and Data Acquisition*) que son utilizados en la supervisión de energía (Qiu *et al.*, 2002). Sin embargo, la información aportada por estos sistemas suele ser bastante escasa y reducida a la visualización de la evolución de las variables de forma individual, lo cual dificulta la extracción de conocimiento por personas no versadas en la materia. Con el fin de reducir la magnitud de este problema, es posible la creación de herramientas basadas en técnicas de supervisión avanzadas que son capaces de extraer el conocimiento implícito en la totalidad de los datos y trasladar dicho conocimiento a la toma de decisiones (Khan, 2001).

Esta extracción de conocimiento facilita el análisis de la evolución del consumo de un edificio, lo cual es útil no solo en el campo de la eficiencia y la negociación de las facturas con la compañía suministradora, sino también para resolver problemas en tiempo real en la gestión de desviaciones, planificación de los consumos y eliminación de consumos anómalos

(Dahhaghchi *et al.*, 1997). Dado que tanto la cantidad de datos como el número de variables a analizar es elevado, se hace necesario el uso de técnicas de aprendizaje automático (*Machine learning*) (Alpaydin, 2004) ya que las relaciones existentes entre los datos son muy complejas como para ser tratadas por algoritmos básicos.

Los algoritmos de aprendizaje automático se pueden emplear para reducir la dimensionalidad de los datos, lo que facilita la visualización y la extracción de información implícita (Kalogirou, 2000). Por tanto, es necesario realizar un estudio de los distintos métodos para reducir la dimensionalidad que existen y ver si su aplicación a perfiles de consumo es útil a la hora de extraer conocimiento acerca del comportamiento del edificio a estudiar. Con el fin de comprobar la eficacia de estos métodos se escoge como campo de experimentación los edificios del campus de la Universidad de León, los cuales presentan comportamientos dispares. Por ello, el estudio de las diferentes dinámicas de consumo va a permitir no solo descubrir ineficiencias en este tipo de edificios, sino que ofrece también la posibilidad de abrir nuevas líneas de investigación para mejorar la eficiencia y potenciar el ahorro. Además, estas técnicas abren la posibilidad a nuevas estrategias basadas en la predicción, no solo porque la información devuelta por estos métodos permite conocer mejor la evolución de las curvas de consumo, sino también por que al reducir los datos representativos permite generar modelos de predicción más precisos de los datos (Barreto, 2007).

1.1.2. Minería de datos

La minería de datos (*Data Mining*) busca la extracción de conocimiento implícito, potencialmente útil que permanece oculto en grandes volúmenes de datos con el fin de encontrar patrones y resumirlos de forma que sean fácilmente entendibles por parte de un usuario (Hand *et al.*, 2001). Se trata de una tarea interdisciplinar del campo de la informática que combina métodos estadísticos, inteligencia artificial, aprendizaje automático, reconocimiento de patrones y gestión de las bases de datos. Cuando se combinan estas técnicas con técnicas de visualización surge la minería de datos visual (*Visual Data Mining*) (Wong, 1999) que permite obtener una transformación de los datos para generar representaciones gráficas en las cuales un ser humano pueda operar de forma visual, razonando con imágenes.

La minería de datos persigue fundamentalmente una exploración de los datos que busca examinar de forma global el conjunto de datos (Tukey, 1977) y descubrir patrones y reglas (*Pattern recognition*). Este objetivo es una parte del aprendizaje automático, cuya función es descubrir automáticamente regularidades en los datos utilizando algoritmos ejecutados en un ordenador (Bishop, 2006). Estos algoritmos buscan encontrar generalizaciones o reducciones de los datos siguiendo estrategias de optimización. El aprendizaje se puede dividir en supervisado, es decir, aquel que requiere especificaciones de las salidas e información conocida del entorno, y no supervisado, donde las salidas del modelo son desconocidas y se intenta representar la estructura interna de los datos que minimiza un criterio de error.

En cuanto a la visualización de los datos, que es el otro objetivo de la minería de datos visual, se busca presentar la mayor cantidad de datos posibles al analista de un modo que agilice su razonamiento y facilite el reconocimiento de anomalías, tendencias, patrones,

estructuras, novedades o correlaciones. También pretende facilitar la comparación entre valores y la búsqueda de datos, la comparación con modelos y el descubrimiento de errores o detalles inesperados (Keim, 2001). Un usuario sin conocimientos específicos puede realizar interpretaciones intuitivas y rápidas de los patrones existentes en los datos (Hinneburg *et al.*, 2003; Ferreira de Oliveira y Levkowitz, 2003). De esta forma se aprovecha el conocimiento extraído para integrar al ser humano en un lazo de control de proceso para controlar la evolución del mismo.

Esta integración del ser humano en los lazos de control permite aprovechar la capacidad de razonamiento de un ser humano para adquirir un conocimiento y aportar una realimentación que corrija el sistema, siendo el que toma las decisiones en última instancia (Keim *et al.*, 2010). Dado que gran parte de la función cognitiva humana tiene lugar de manera visual (Ware, 2008), una visualización generada por medio de una técnica de proyección, que reduce la dimensionalidad de los datos, permite aprovechar en mayor medida la capacidad de procesamiento humana.

El objetivo final de la visualización es lograr un equilibrio entre la utilidad de las visualizaciones y su vistosidad con el fin de que no solo muestren información útil de forma sencilla, sino que también sean capaces de estimular la atención del usuario (Tufte, 1983). Por ello, cada vez se hace más necesario el desarrollo de herramientas de visualización interactivas en las que no solo se muestre la información, sino que también se ofrezca al usuario la posibilidad de modificar dichas visualizaciones con el fin de lograr una extracción de información en base a las transiciones que se puedan dar entre las diferentes visualizaciones.

Se han propuesto gran cantidad de técnicas de visualización de conjuntos de datos multidimensionales que utilizan diversos métodos (Keim, 2002):

- Transformaciones geométricas donde un punto de un espacio n -dimensional se representa por medio de una polilínea.
- Técnicas orientadas a píxeles en las que los valores se codifican mediante colores.
- Múltiples gráficas simultáneas que permiten una inspección visual eficaz, ya que es posible interpretarlas de forma similar.
- Visualizaciones icónicas que hacen uso de marcadores visuales parametrizables.

El principal problema de estas técnicas es que no reducen por sí mismas ni la dimensión ni la cantidad de datos (Keim, 1996). Por ello es necesario utilizar técnicas de **reducción de la dimensionalidad** que proyecten de forma efectiva los datos en una dimensión más baja. No obstante, la mayor parte de ellas no han sido específicamente diseñadas para la visualización. De igual manera, es necesario simplificar el conjunto de datos realizando una compresión de los mismos. Para ello, se hace uso de algoritmos de **cuantización de vectores** (VQ). Estos algoritmos proporcionan una representación eficiente y compacta de los datos, proporcionando un conjunto de vectores prototipo, que minimizan alguna medida de distorsión. Esta simplificación del conjunto de datos a información estrictamente relevante hace posible la abstracción que se requiere para su tratamiento.

La reducción de la dimensionalidad (*Dimensionality Reduction*, DR) tiene como objetivo encontrar un conjunto variables que generen un espacio de baja dimensión visualizable y que conserven, en la medida de lo posible, la información del conjunto de datos original (Carreira-Perpiñán, 1996). La reducción de la dimensionalidad busca, por un lado, transformar los datos de entrada de tal forma que se generen representaciones que faciliten el reconocimiento de patrones en los datos y, por otro, representar dichos datos en un sistema de coordenadas visualizable. La reducción del número de variables es posible gracias a que la dimensión intrínseca o latente de los datos suele ser menor que la real, debido a la existencia de variables correlacionadas y con ruido e información redundante (Lee y Verleysen, 2007). Algoritmos característicos como el *MultiDimensional Scaling* (MDS), *Isometric Mapping Method* o *Curvilinear Component Analysis* (CCA) se basan en la conservación de la distancia entre los datos originales y los datos proyectados. Por otro lado, la relación entre el espacio de entrada y el espacio de salida también puede ser de topología, conservándose entonces la “forma” de los datos. Algoritmos destacables en este sentido son el *Locally Linear Embedding* (LLE) y el *Laplacian Eigenmap* (LE). La aplicación de un tipo u otro de algoritmos depende de la complejidad de los datos y de la eficacia que se quiera lograr con las proyecciones.

La simplificación de los datos, por medio de una compresión, hace más sencillo y rápido el procesamiento de los mismos. Las técnicas de cuantización de vectores (*Vector Quantization*, VQ) tienen como objetivo modelar la función de densidad de probabilidad de los datos (Gray y Neuhoﬀ, 1998) y encontrar los datos más representativos del espacio de entrada. Esto permite transformar un gran conjunto de datos de alta dimensión en un conjunto pequeño de datos, que se denominan prototipo. Las técnicas de cuantización se pueden considerar similares a las técnicas de agrupamiento o *clustering*, ya que ambas son capaces de dividir los datos de entrada en una serie de grupos representativos, minimizando la distorsión, y que pueden ser representados por medio de un vector que constituye el centroide de dicho grupo.

Entre todas las técnicas utilizadas en la reducción de la dimensionalidad y cuantización de vectores, destaca el mapa autoorganizado (*Self-Organizing Map*, SOM). El SOM es una red neuronal no supervisada, basada en aprendizaje competitivo que implementa un mapeado no lineal de un espacio de entrada de datos de alta dimensión en un espacio de salida de baja dimensión. Las neuronas que constituyen el SOM forman una malla, normalmente de 2 dimensiones, que esta topológicamente ordenada. (Kaski, 1997; Vesanto, 2002).

Todas estas técnicas de minería de datos pueden ser aplicadas a variables de consumo energético con el fin de extraer información de los datos y así poder conocer el comportamiento dinámico de estas variables. Este análisis permite no solo conocer la dependencia horaria de las variables, sino también analizar la correlación entre la variable de consumo y otras variables externas al sistema. El conocimiento del comportamiento dinámico permite comparar edificios entre sí y por tanto agruparlos, detectar desviaciones y el porqué de las mismas.

1.1.3. Predicción

Existe una gran cantidad de técnicas que permiten la estimación de los valores futuros de una determinada variable. Algunas de las técnicas más utilizadas están basadas en series temporales y tratan de hacer una estimación de la evolución de la variable a predecir, en función de la evolución pasada (Weigend y Gershenfeld, 1993). Muchas de estas técnicas utilizan la información obtenida por medio de la minería de datos con el fin de obtener modelos de predicción que sean aplicados al problema de las series temporales. A lo largo de la bibliografía se han utilizado multitud de técnicas que utilizan redes neuronales tales como MLP (*MultiLayer Perceptron*) y RBF (*Radial Basis Function*) (Leung *et al.*, 2001; Fan *et al.*, 2009), técnicas de regresión como ARIMA (*AutoRegressive Integrated Moving Average*), SVR (*Support Vector Regression*) (Tiao y Tsay, 1994; Cao, 2003), el SOM, que constituye una entidad propia gracias a la gran cantidad de modificaciones que son utilizadas para predicción (Barreto, 2007), redes *fuzzy*, sistemas expertos y métodos híbridos que combinan varios métodos (Fan *et al.*, 2006), por citar algunos.

La predicción de las variables de consumo energético permite estimar la evolución futura de estas variables con el fin de poder anticiparse a posibles picos y conocer el consumo a lo largo del día siguiente. En el caso de consumos eléctricos, la predicción es una herramienta imprescindible a la hora de programar la producción de energía por parte de las centrales eléctricas. La energía eléctrica sufre el problema de que no se puede almacenar de forma eficaz, por lo que toda la energía producida ha de consumirse. Por esta razón, un conocimiento lo más exacto posible de la evolución del consumo permite reducir al máximo las pérdidas por un exceso de energía producida.

Las series temporales han sido aplicadas con éxito en la predicción de grandes consumos energéticos, como son los consumos de un país (Lendasse *et al.*, 2002) y, a menor escala, islas (Sideratos *et al.*, 2011) o incluso ciudades (Fan *et al.*, 2009). Estas estimaciones buscan sobre todo lograr un equilibrio entre la producción y el consumo de energía. También existen aplicaciones de las series temporales utilizadas en la predicción de la evolución del precio de la energía eléctrica (Martínez Álvarez *et al.*, 2011; Chen *et al.*, 2008) que utilizan este tipo de técnicas. El estudio de ciertas variables ambientales que pueden condicionar el consumo energético de los edificios resulta interesante a la hora de mejorar la precisión de las predicciones (Gupta y Yamada, 1972; Fan *et al.*, 2006). Por lo tanto, el análisis de la dependencia con las variables ambientales va a permitir incorporar esta información a los modelos con el fin de mejorar los resultados de predicción.

Algoritmos efectivos en la predicción de grandes consumos no tienen por qué generar los mismos resultados en el caso de edificios, ya que se modifica la magnitud del problema. Son pocas las aplicaciones de técnicas de predicción a la estimación del consumo de un edificio (González y Zamarreño, 2005; Cao, 2003) y normalmente se suelen aplicar a edificios de gran consumo debido a que, cuanto menor sea el consumo, más difícil será la predicción y mayor el error. La precisión de estos métodos depende en gran medida de la naturaleza de la serie temporal, de forma que determinar si la serie es determinista o estocástica, lineal o no, estacionaria, continua o discreta, ayuda en gran medida a escoger el método de predicción y a mejorar los resultados (Fay *et al.*, 2003; Murray *et al.*, 2000). Por ello,

el conocimiento extraído del análisis de los perfiles de consumo de los edificios mediante la minería de datos resulta de utilidad, ya que va a permitir seleccionar los regresores, que incorporan las variables que tienen una mayor influencia en el consumo futuro, para generar los modelos de predicción.

La aplicación de algoritmos de predicción ya conocidos para la estimación del consumo en edificios del sector terciario resulta interesante por tres razones principales:

- Una predicción a largo plazo proporciona una posición ventajosa con respecto a la compañía suministradora al poder conocer el consumo estimado para un año y poder así negociar los términos del contrato.
- Conocer con anterioridad los posibles picos de consumo y poder programar la entrada de grandes equipos, como pueden ser climatizadores, y evitar así un exceso de potencia que será registrada en el maxímetro.
- Una predicción fiable a un día va a permitir no solo hacer una programación del consumo, sino que también abre la posibilidad de acudir directamente al mercado de subasta con el fin de ahorrar los costes de intermediarios.

1.2. Objetivos de la tesis

Los objetivos planteados se centran en:

1. Implantar una plataforma de comunicaciones que permita la adquisición, almacenamiento y procesamiento de datos relacionados con el consumo energético de edificios públicos. Esta plataforma tratará de simplificar las estructuras existentes aumentando la seguridad y eficiencia de las mismas.
2. Modelar el comportamiento de los edificios públicos en base a sus perfiles de consumo energético mediante el uso de técnicas de minería de datos. La obtención de estos modelos y el estudio de su dependencia con variables ambientales externas como la temperatura, humedad, radiación solar, horario de actividad, etc. permitirán determinar el patrón de consumo de un edificio y comparar edificios entre sí.
3. Desarrollar técnicas para predecir los perfiles de consumo energético en los edificios. En este sentido, se estudiarán técnicas de predicción a corto y largo plazo.
4. Analizar la información de facturación suministrada por la compañía y mejorar la visualización de dicha información mediante el uso de técnicas de reducción de la dimensionalidad que proporcionen un análisis intuitivo y sencillo.
5. Desarrollar herramientas que recojan los resultados obtenidos en el modelado, predicción y análisis, e incorporarlas en la plataforma de supervisión energética.

1.3. Estructura de la tesis

Esta tesis se estructura de la siguiente forma:

- En el capítulo 2 se hace una revisión de las principales técnicas de reducción de la dimensionalidad y cuantización de vectores que tienen por objetivo el análisis de grandes volúmenes de datos.
- En el capítulo 3 se revisan las principales técnicas existentes para la predicción de series temporales que se derivan de las técnicas de reducción de la dimensión y cuantización.
- En el capítulo 4 se explica la arquitectura del sistema de adquisición, almacenamiento y procesamiento diseñado para llevar a cabo la supervisión del consumo energético de los edificios de la Universidad de León.
- En el capítulo 5 se propone la metodología basada en técnicas de reducción de la dimensionalidad para la visualización y comparación de los perfiles de consumo entre los edificios. También se proponen las técnicas para agrupar los edificios, seleccionar los regresores y los métodos para predicción a corto y largo plazo.
- En el capítulo 6 se explican los experimentos realizados para verificar la metodología adoptada y se muestran los resultados obtenidos en esta tesis.
- En el capítulo 7 se exponen las conclusiones extraídas en esta tesis, se enumeran las aportaciones científicas y técnicas y se indican posibles líneas futuras de investigación.

Reducción de la dimensionalidad

Existen multitud de métodos que permiten transformar un conjunto de datos originales a analizar en un conjunto de características, más reducido, pero que sigue conservando, de forma significativa, la información que poseen los datos de partida. Estos métodos se centran en preservar un aspecto esencial de los datos en función del objetivo que se quiera lograr, aunque frecuentemente el objetivo principal es la visualización.

En la bibliografía existen, por un lado, técnicas que permiten reducir el número de variables que conservan en gran medida la información de los datos de partida, dando lugar a las técnicas de reducción de la dimensionalidad, bien sea conservando la distancia presente entre los datos o bien preservando la topología. Por otro lado, se tienen técnicas que permiten obtener las muestras representativas, reduciendo así el número de datos de partida, que se agrupan bajo el nombre de técnicas de cuantización o *clustering*.

2.1. Introducción a la reducción de la dimensionalidad

El mundo es multidimensional; sin embargo, los algoritmos que analizan los datos que el mundo proporciona padecen lo que se conoce como **maldición de la dimensionalidad** (Bellman, 1961) que implica que la complejidad y el número de datos de entrenamiento necesarios dependa directamente de la dimensión de entrada.

La reducción de la dimensionalidad permite, en algunos casos, reducir la complejidad de los datos para que estos puedan ser manejados por otros algoritmos; en otros casos, obtener información visualizable con el objetivo de extraer conocimiento.

Algunos de los campos de aplicación de estas técnicas son:

- Procesado de datos de redes de sensores.
- Procesamiento de imágenes.
- Análisis de datos multivariable.
- Minería de datos.

Las técnicas de reducción de la dimensionalidad se pueden aplicar principalmente a dos tipos de datos. Por un lado, **datos espaciales** que están constituidos por conjuntos de

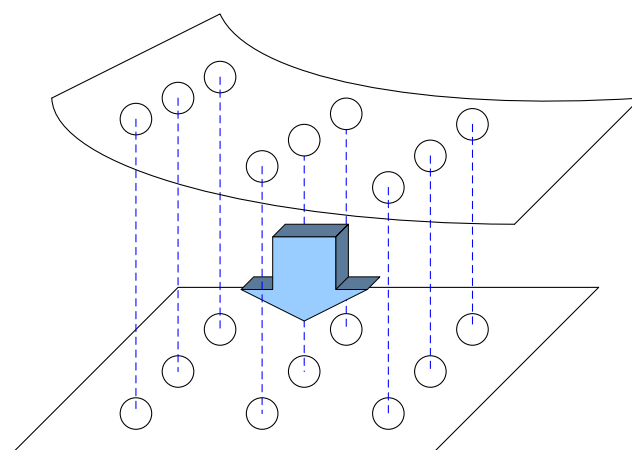


Figura 2.1: Reducción de la dimensionalidad.

variables que se pueden asemejar a coordenadas espaciales, lo cual hace que la visualización, cuando el número de variables es mayor que tres, sea bastante compleja. Por otro lado, los **datos temporales**, que incorporan al tipo de datos anteriores información temporal, añadiendo a la visualización su evolución con el tiempo.

Se puede reducir la dimensionalidad de los datos porque generalmente suelen contener información redundante, correlaciones o variables que pueden ser desechadas. Por eso, el número de variables independientes que pueden describir el sistema, que es lo que se conoce como **dimensión intrínseca**, es menor que su dimensión nominal (Carreira-Perpiñán, 2001). El cálculo de esta dimensión es el principal problema, ya que la eficacia de la reducción depende en gran medida de este parámetro. Existen diversos métodos que permiten inferir valores para este parámetro. No obstante, para los casos de visualización se requiere que la dimensión sea 2 ó 3, ya que son dimensiones visualizables.

Algunos conceptos involucrados en la reducción de la dimensión son:

- **Espacio de entrada.** Es un espacio de p -dimensiones, \mathbb{R}^p , que está formado por los datos crudos en los que cada dimensión p está formada por una de las variables.
- **Espacio de salida.** Es un espacio de d -dimensiones, \mathbb{R}^d , siendo d , idealmente, la dimensión intrínseca del conjunto de datos y $p > d$, con lo que se produce la reducción de la dimensionalidad.
- **Proyección.** Es el paso del espacio p -dimensional o de dimensión elevada al espacio d -dimensional o de baja dimensión para la generación de visualizaciones (ver figura 2.1). A lo largo de este trabajo se emplea este término para referirse a transformaciones lineales, no lineales y *embedding*, ya que lo que se busca en estas transformaciones es la preservación de características geométricas o que pueden interpretarse como tales.
- **Manifold o variedad topológica.** Es un subespacio topológico \mathfrak{R} . Se refiere a un espacio topológico que se puede considerar localmente euclídeo, lo que significa que

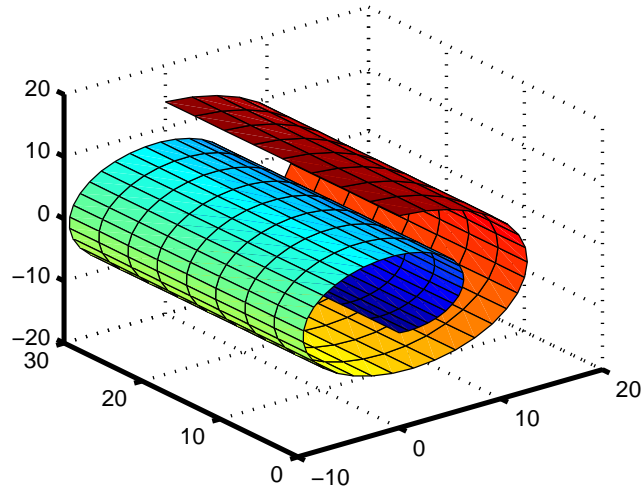


Figura 2.2: *Manifold Swiss-Roll*.

alrededor de cada punto de \mathcal{R} , existe una vecindad que es topológicamente idéntica a la vecindad del punto en el espacio de entrada. Se puede decir que, en general, aquellos objetos que son casi “planos” a pequeña escala forman un *manifold*. Por ejemplo, la tierra es esférica pero parece plana a escala humana (Lee, 2000).

En la figura 2.2 se muestra un ejemplo de *manifold* típico en la literatura (Tenenbaum *et al.*, 2000), que se conoce como *Swiss-Roll*. Este *manifold* consiste en un conjunto de puntos discretos situados sobre una capa que se enrolla sobre sí misma. Estos puntos tienen una dimensión intrínseca de 2, incrustada dentro de un espacio de 3 dimensiones. Para cada método de reducción de la dimensionalidad se analizarán sus resultados utilizando este conjunto de datos.

Una importante utilidad de las técnicas de reducción de la dimensionalidad es la visualización de los datos de una manera entendible por el ser humano, ya que estas visualizaciones permiten la detección de patrones de forma rápida y eficiente por medio de la capacidad del sistema visual humano (Tufte, 1983). El uso de algoritmos de reducción de la dimensionalidad permite la transformación de los datos, sin una gran pérdida de información, en una representación entendible por un usuario y si, además, se las dota de mecanismos de interacción, se facilita la exploración y el análisis (Fray, 2004).

Existe un elevado número de técnicas para lograr la reducción de la dimensionalidad, que se pueden agrupar fundamentalmente en dos tipos en función de que traten de preservar la distancia de los datos de entrada o la topología de los mismos (Lee y Verleysen, 2007). Uno de los métodos más antiguos y mejor conocidos es el *Principal Component Analysis* (PCA), el cual fue introducido por Pearson (1901) y desarrollado posteriormente por Hotelling (1933). Este algoritmo genera una colección de variables \mathbf{y}_i , combinación lineal de las variables observadas \mathbf{x}_i , que tienen la propiedad de poseer una varianza máxima y estar incorreladas. Estas propiedades de las componentes principales permiten explicar, mediante unas pocas

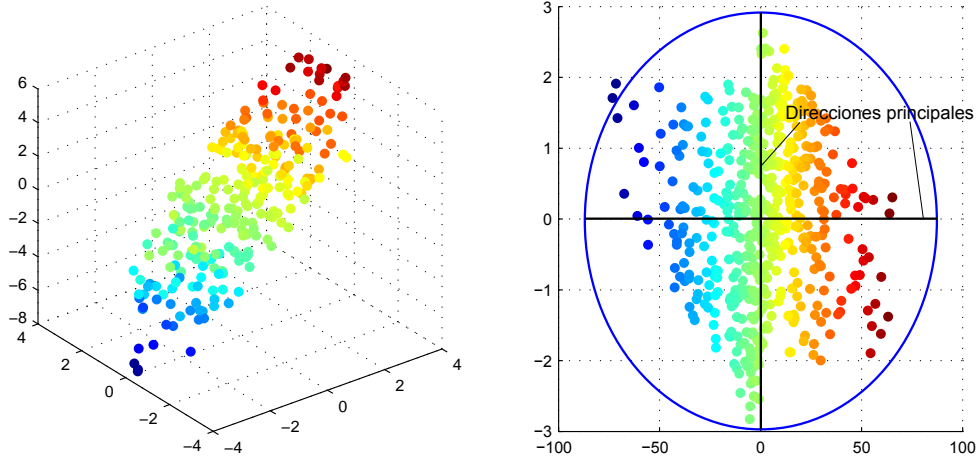


Figura 2.3: Ejemplo de proyección PCA.

variables, la mayor parte de la información que contienen las variables observadas (ver figura 2.3).

Para calcular el PCA, se obtiene la matriz \mathbf{U} , que contiene los vectores propios correspondientes a los valores propios mayores de la matriz de covarianzas $\mathbf{\Sigma}$, llamados componentes principales:

$$\Sigma = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (2.1)$$

$$\mathbf{y}_i = \mathbf{U}^T \mathbf{x}_i$$

donde $\bar{\mathbf{x}}$ es la media de los vectores y $\mathbf{\Lambda}$ es una matriz diagonal que contiene los valores propios. Puesto que $\mathbf{\Lambda}$ es diagonal, las variables de salida están incorreladas, por lo que la información que aportan no es redundante.

Dada la sencillez de este método, ha sido ampliamente utilizado ya que produce proyecciones que son fáciles de interpretar. Su principal defecto aparece cuando los datos no son lineales ya que tan solo representa una transformación lineal ortogonal de los datos. Para paliar este problema se han propuesto diversas generalizaciones, como son las redes neuronales autoasociativas, las curvas principales y el *kernel* PCA.

2.2. Reducción de la dimensionalidad por preservación de la distancia

Los métodos más simples que permiten reducir la dimensionalidad, aparte del PCA, son aquellos que tratan de preservar la distancia en los puntos proyectados con respecto

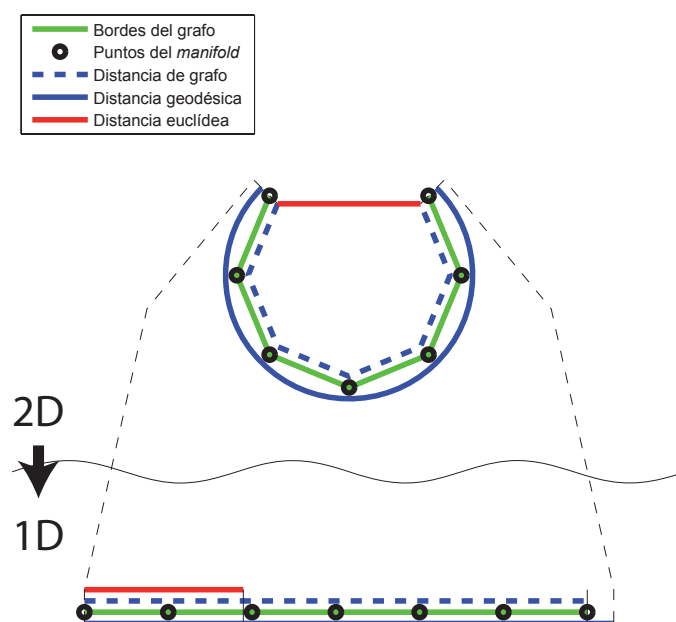


Figura 2.4: Comparación de los distintos tipos de distancia. (Imagen de Lee y Verleysen (2007))

a los puntos en el espacio de entrada (Venna, 2007). Estos métodos tratan de mantener propiedades geométricas de los datos tales como la forma global o las relaciones de vecindad. Dado que para casos no lineales las distancias no pueden ser preservadas perfectamente, los distintos métodos tratan de superar esta dificultad.

A la hora de calcular las distancias, se pueden utilizar dos tipos de distancias (Lee y Verleysen, 2007):

- **Distancia espacial.** Las distancias espaciales, como la euclídea, son las más intuitivas y es la forma natural de medir las distancias. Estas distancias se calculan teniendo en cuenta únicamente las coordenadas de los dos puntos cuya distancia se quiere medir. Los métodos basados en este tipo de distancia tienen la ventaja de que son muy simples aunque no siempre obtienen los mejores resultados.
- **Distancia geodésica.** Es la distancia que hay entre dos puntos a lo largo del *manifold* teniendo en cuenta la topología, de tal forma que la curvatura de los datos es analizada a la hora de calcular las distancias. La distancia geodésica se puede interpretar como una vía de tren en la cual los trenes están forzados a seguir la vía que en este caso sería el *manifold*. Calcular la distancia geodésica es complejo por lo que normalmente lo que se calcula es la **distancia de grafo**. Esta distancia es una suma de las distancias euclídeas entre los puntos del grafo generado para interconectarlos, siendo este grafo generado considerando la topología del *manifold*. Para ello, se suelen usar algoritmos como el de Dijkstra (Dijkstra, 1959).

En la figura 2.4 se muestra la diferencia existente entre los tipos de distancia.

2.2.1. Métodos de escalado multidimensional

Los métodos *Multidimensional Scaling* (MDS) son un conjunto de métodos más que un procedimiento definido. Estos métodos tienen como objetivo generar un conjunto de puntos en un espacio de baja dimensión cuyas distancias mutuas se asemejen lo más posible a las del espacio de origen. El punto de partida de los algoritmos MDS es una matriz formada por las distancias mutuas entre los vectores $\{\mathbf{x}_k\}_{k=1,\dots,N}$ del espacio de entrada, que en el caso de una distancia euclídea son:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\langle (\mathbf{x}_i - \mathbf{x}_j), (\mathbf{x}_i - \mathbf{x}_j) \rangle}, \quad (2.2)$$

donde $\langle \cdot, \cdot \rangle$ denota el producto escalar. La distancia, en principio, no tiene por qué restringirse a la distancia euclídea pudiéndose utilizar otras normas.

El objetivo de estos métodos es generar una colección de puntos $\{\mathbf{y}_k\}_{k=1,\dots,N}$ de un espacio de dimensión inferior, de manera que su matriz de distancias mutuas $d(\mathbf{y}_i, \mathbf{y}_j)$ se asemeje lo más posible a la del espacio de entrada. La matriz de distancias mutuas de un conjunto de puntos define su topología, es decir, la forma en que estos están dispuestos.

La determinación de las proyecciones en estos algoritmos se lleva a cabo a través de la minimización de funciones de coste, que penalizan disparidades entre las matrices de distancias mutuas, siendo precisamente estas funciones de coste las que marcan las diferencias entre los distintos algoritmos. La forma más básica de este tipo de algoritmos es aquella cuya función de coste E es:

$$E_{MDS} = \sum_{i \neq j} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2. \quad (2.3)$$

Esta función constituye el **MDS métrico** (Kruskal y Wish, 1978), que pondera por igual todas las disparidades entre distancias mutuas que puedan darse entre los datos de entrada y los de salida. Constituye un algoritmo de tipo lineal, similar al PCA, que se resuelve como un problema de autovalores. Sin embargo, conserva mejor las distancias largas, lo que preserva mejor la estructura global de los datos.

Al ser similar al PCA, presenta las mismas ventajas y desventajas. Por comparación, el MDS es más flexible que el PCA pero requiere una mayor cantidad de memoria en el procesamiento.

El *Isometric Feature Mapping* (Isomap) (Tenenbaum *et al.*, 2000) es una variante del escalado multidimensional que utiliza distancias geodésicas. El resultado, que busca una optimización global, se puede obtener aplicando MDS lineal a la matriz de distancias mínimas. Este método también puede considerarse un método de aprendizaje de la variedad (*manifold learning*), ya que considera la forma de los datos en el espacio de entrada. Además proporciona una estimación de la dimensión de la misma por medio del número de autovalores distintos de cero encontrados por el algoritmo. Bajo ciertas condiciones, se garantiza que el Isomap converge a una parametrización correcta de la variedad. La principal ventaja del Isomap es que, a diferencia del MDS o del PCA, puede tratar con una gran cantidad de datos, aunque la escasez de datos de entrenamiento, la dependencia de la vecindad

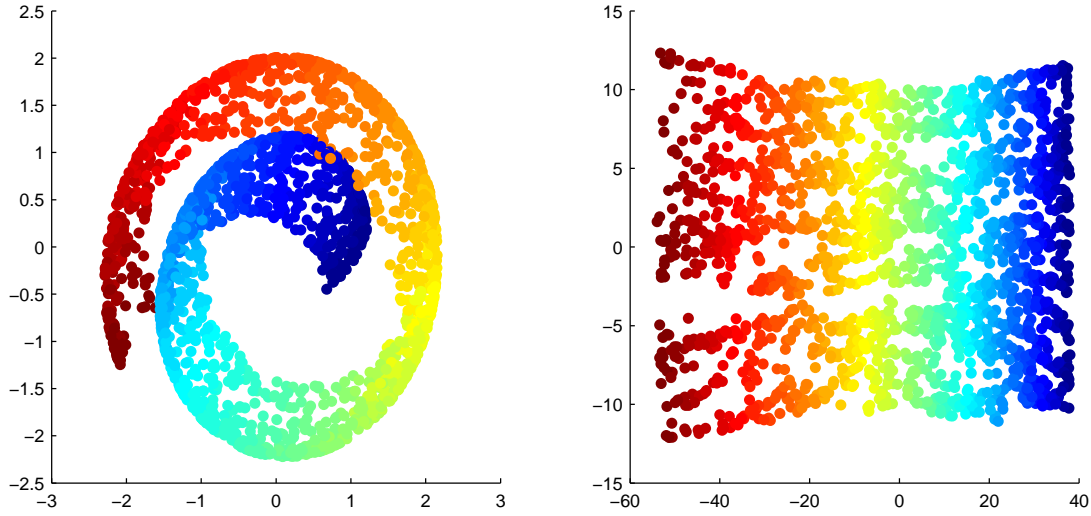


Figura 2.5: Proyección del *Swiss-Roll* con MDS e Isomap.

y la complejidad del algoritmo dificultan el correcto aprendizaje de la variedad topológica. En la figura 2.5 se muestra la proyección del *Swiss-Roll* por ambos métodos.

Algoritmo Isomap

1. Construir los grafos con los k vecinos más próximos para todos los puntos.
2. Calcular las rutas más cortas utilizando el algoritmo de Dijkstra o de Floyd.
3. Realizar una proyección MDS al espacio euclídeo de baja dimensionalidad.

2.2.2. Proyección de Sammon

El *Nonlinear Mapping* (NLM) (Sammon, Jr., 1969) tiene como propósito reducir la dimensionalidad de un número finito de puntos. Se puede considerar como un tipo de algoritmo MDS que utiliza una función de coste normalizada en función de las distancias en el espacio de entrada:

$$E_{NLM} = \frac{1}{\sum_{\substack{i=1 \\ i < j}}^N d(\mathbf{x}_i, \mathbf{x}_j)} \sum_{\substack{i=1 \\ i < j}}^N \frac{(d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2}{d(\mathbf{x}_i, \mathbf{x}_j)}. \quad (2.4)$$

Al igual que en el caso del MDS, $d(i, j)$ representa las distancias euclídeas entre los puntos i y j en los espacios de entrada y salida. La proyección de Sammon intenta preservar las distancias relativas entre los datos, dando un mayor peso a las distancias cortas en el espacio de entrada. De esta forma, se preservan las distancias locales y se busca que la estructura de los datos en el espacio de baja dimensión sea lo más similar posible a la de los datos de entrada.

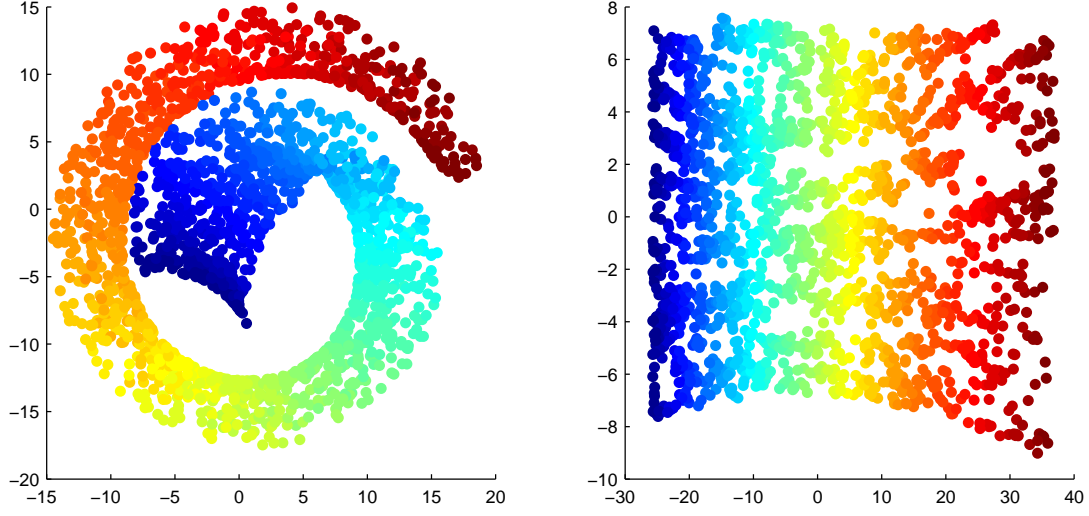


Figura 2.6: Proyección del *Swiss-Roll* con NLM y GNLM.

El proceso de minimización de la función de coste se lleva a cabo de forma iterativa, aplicando técnicas estándar como es una variación del método de Newton. Esta modificación del método de Newton se basa en la siguiente regla de actualización, que calcula las coordenadas del espacio de salida \mathbf{y} :

$$\Delta \mathbf{y}_i = \alpha \frac{\left(\frac{\partial E}{\partial \mathbf{y}_i} \right)}{\left(\frac{\partial^2 E}{\partial \mathbf{y}_i^2} \right)}. \quad (2.5)$$

Sammon denomina en su artículo al factor α como factor mágico y recomienda un valor entre 0.3 y 0.4.

La ventaja de este método es que computacionalmente es muy simple y obtiene buenos resultados para conjuntos de datos no lineales, siempre y cuando no sean muy complejos. Su principal inconveniente es que, en algunos casos, el proceso de optimización puede quedarse bloqueado en un mínimo local.

Existe una modificación del NLM que usa como función de coste distancias geodésicas en vez de distancias euclídeas, en el espacio de entrada. Este método se conoce como ***Geodesic Nonlinear Mapping (GNLM)*** (Yang, 2004). En este caso la función de coste a minimizar es muy parecida a la del NLM:

$$E_{GNLM} = \frac{1}{\sum_{\substack{i=1 \\ i < j}}^N \delta(\mathbf{x}_i, \mathbf{x}_j)} \sum_{\substack{i=1 \\ i < j}}^N \frac{(\delta(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2}{\delta(\mathbf{x}_i, \mathbf{x}_j)}. \quad (2.6)$$

En esta ecuación, $\delta(\mathbf{x}_i, \mathbf{x}_j)$ se refiere a la distancia geodésica en el espacio de entrada entre dos puntos. Por lo tanto, como paso previo en este método, hay que calcular el grafo en el espacio de entrada.

Algoritmo Sammon

1. Calcular las distancias mutuas $d(\mathbf{x}_i, \mathbf{x}_j)$ en el espacio de entrada.
2. Inicializar las coordenadas \mathbf{y}_i de los datos en el espacio de salida \mathbb{R}^d . Se puede emplear el método PCA o bien, proporcionar valores aleatorios.
3. Calcular la variación de las coordenadas de los datos del espacio de salida \mathbf{y}_i con la ecuación 2.5.
4. Actualizar las coordenadas del espacio de salida.
5. Volver al paso 3 hasta que el valor de la función de error E_{NLM} no disminuya.

La principal ventaja de este método es que la proyección es más precisa para datos que presentan un gran plegamiento. En la figura 2.6 se puede ver la diferencia entre utilizar distancias euclídeas y geodésicas.

2.2.3. Análisis de componentes curvilíneas

El *Curvilinear Component Analysis* (CCA) fue propuesto por Demartines y Hérault (1997) como una mejora de los mapas autoorganizados de Kohonen (Kohonen, 1982). A diferencia de este método, que se verá más adelante, el CCA pertenece a los métodos de preservación de la distancia, estando más relacionado con la proyección de Sammon.

Este método minimiza una función de coste basada en distancias entre puntos, definida por la ecuación 2.7, tanto en el espacio de entrada como en el de salida.

$$E_{CCA} = \frac{1}{2} \sum_i \sum_{j \neq i} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 F(d(\mathbf{y}_i, \mathbf{y}_j), \lambda_y), \quad (2.7)$$

donde $d(\mathbf{x}_i, \mathbf{x}_j)$ y $d(\mathbf{y}_i, \mathbf{y}_j)$ son los elementos de las matrices de distancias mutuas entre puntos del espacio de entrada y de salida. Generalmente, F es una función acotada, monótona decreciente, tal y como la que se indica en la siguiente ecuación:

$$F(d(\mathbf{y}_i, \mathbf{y}_j), \lambda_y) = \begin{cases} 1 & \text{si } d(\mathbf{y}_i, \mathbf{y}_j) \leq \lambda_y \\ 0 & \text{si } d(\mathbf{y}_i, \mathbf{y}_j) > \lambda_y \end{cases} \quad (2.8)$$

El parámetro λ_y decrece a lo largo del entrenamiento y permite controlar la escala a la que trabaja el algoritmo.

CCA es un algoritmo de menor complejidad computacional, puesto que en cada etapa se fija un punto \mathbf{y}_i y se adapta el resto, \mathbf{y}_j . De este modo, en cada ciclo de adaptación no es necesario calcular $N(N-1)/2$ distancias, sino solamente las distancias desde el dato i a los demás. Con una $F(d(\mathbf{y}_i, \mathbf{y}_j), \lambda_y)$ discreta, se obtiene la siguiente regla de actualización de los vectores \mathbf{y}_j

$$\Delta \mathbf{y}_j = \alpha(t) F(d(\mathbf{y}_i, \mathbf{y}_j), \lambda_y) (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j)) \frac{\mathbf{y}_j - \mathbf{y}_i}{d(\mathbf{y}_i, \mathbf{y}_j)} \quad \forall j \neq i, \quad (2.9)$$

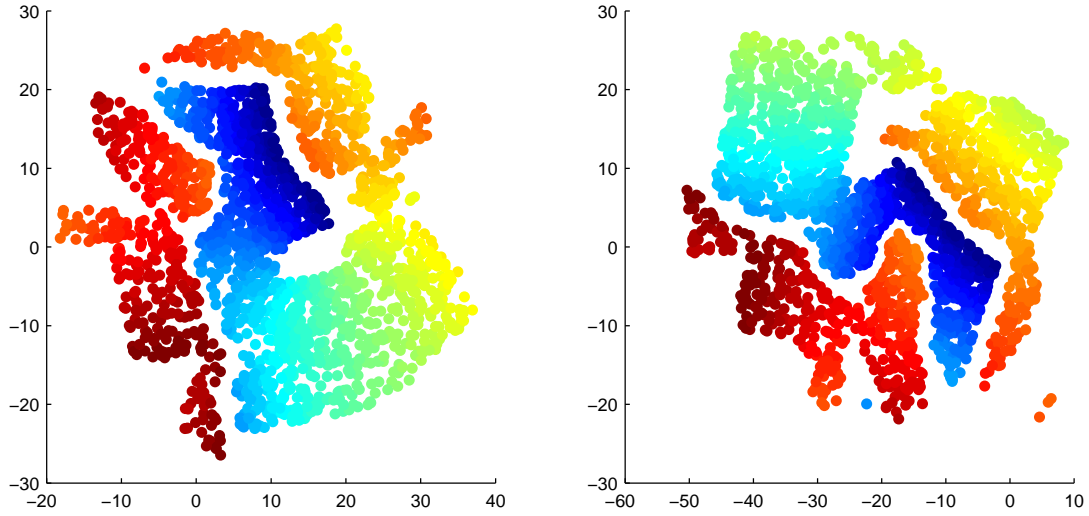


Figura 2.7: Proyección del *Swiss-Roll* con CCA y CDA.

donde $\alpha(t)$ es una función gradiente que decrece con el tiempo para evitar grandes modificaciones en las etapas finales.

Algoritmo CCA

1. Calcular las distancias euclídeas $d(\mathbf{x}_i, \mathbf{x}_j)$ en el espacio de datos p -dimensional.
2. Inicializar las coordenadas del espacio de baja dimensión d -dimensional de todos los puntos \mathbf{y}_i , bien sea de forma aleatoria o mediante una PCA.
3. Actualizar el valor de α y λ_y acorde con la época de entrenamiento.
4. Seleccionar un punto \mathbf{y}_i y actualizar los otros valores de acuerdo a la ecuación 2.9.
5. Repetir el paso 4 para todos los puntos.
6. Incrementar la época de entrenamiento y si no converge volver al paso 4.

Una de las principales ventajas del CCA es que permite la proyección de nuevos datos sin necesidad de entrenar de nuevo. Esta proyección considera los prototipos como puntos fijos y se aplica la regla de actualización (ecuación 2.9) con el fin de mover el nuevo punto a su posición correspondiente. Otra ventaja de este método es que trabaja mejor con estructuras de datos que presentan una elevada curvatura en el espacio de entrada.

Existe una modificación destacable de este algoritmo, denominada **análisis de distancias curvilíneas** (*Curvilinear Distance Analysis* (CDA)) (Lee *et al.*, 2000) que utiliza distancias geodésicas en lugar de euclídeas. Esto lo asemeja más a los mapas de Kohonen al

establecer mejores conexiones con las neuronas vecinas, pero a diferencia de este, la malla no es regular. El CDA mejora el algoritmo anterior, en el sentido de que las distancias son más fiables, pero por el contrario es más difícil calcular la proyección de nuevos puntos y la visualización no es sobre una malla fija y compacta. En la figura 2.7 se muestra la diferencia de proyección entre el CCA y el CDA

2.3. Reducción por preservación de la topología

Estos métodos no tratan de preservar la distancia entre los puntos a la hora de realizar la proyección, sino que tratan de mantener la topología de los datos. Suelen ser más eficaces para realizar la proyección que los basados en distancias, pero también son más complejos. Mientras que los métodos de preservación de distancia se basan únicamente en el cálculo de las distancias, los de preservación de la topología calculan cuáles son los vecinos y tratan de conseguir en la proyección que los vecinos en el espacio de entrada lo sigan siendo en el de salida. El mayor problema en estos métodos es deducir la topología de los datos de entrada.

En función del tipo de topología que utilicen, se pueden diferenciar dos métodos (Lee y Verleysen, 2007):

- **Malla adaptable.** Se utiliza la información contenida en los datos para establecer la topología de los datos y calcular la forma de la proyección acorde a la misma. La proyección no está acotada de ninguna forma y puede adaptarse para capturar la forma del *manifold*.
- **Malla fija.** Se considera siempre el mismo tipo de topología, para lo que se define una malla en el espacio de salida fija y los datos se adaptan a esta malla durante la reducción.

2.3.1. Locally Linear Embedding

El **Locally Linear Embedding (LLE)** (Roweis y Saul, 2000) utiliza *conformal mapping* que es una transformación que trata de preservar los ángulos locales, lo cual está relacionado con las distancias locales. La idea es reemplazar cada punto de los datos por una combinación lineal de los vecinos. De esta manera la geometría del *manifold* puede ser representada por coeficientes lineales que reconstruyen cada dato a través de los vecinos. El error de reconstrucción puede medirse por una función de coste cuadrático:

$$\varepsilon(\mathbf{W}) = \sum_i \left\| \mathbf{x}_i - \sum_j w_{ij} \mathbf{x}_j \right\|^2. \quad (2.10)$$

La matriz \mathbf{W} almacena los pesos de los vecinos para la reconstrucción de \mathbf{x}_i . Para calcular los coeficientes w_{ij} se consideran las siguientes restricciones:

- $w_{ij} = 0$ si el dato \mathbf{x}_j no es vecino de \mathbf{x}_i .
- Las filas de la matriz suman 1 $\sum_j w_{ij} = 1$.

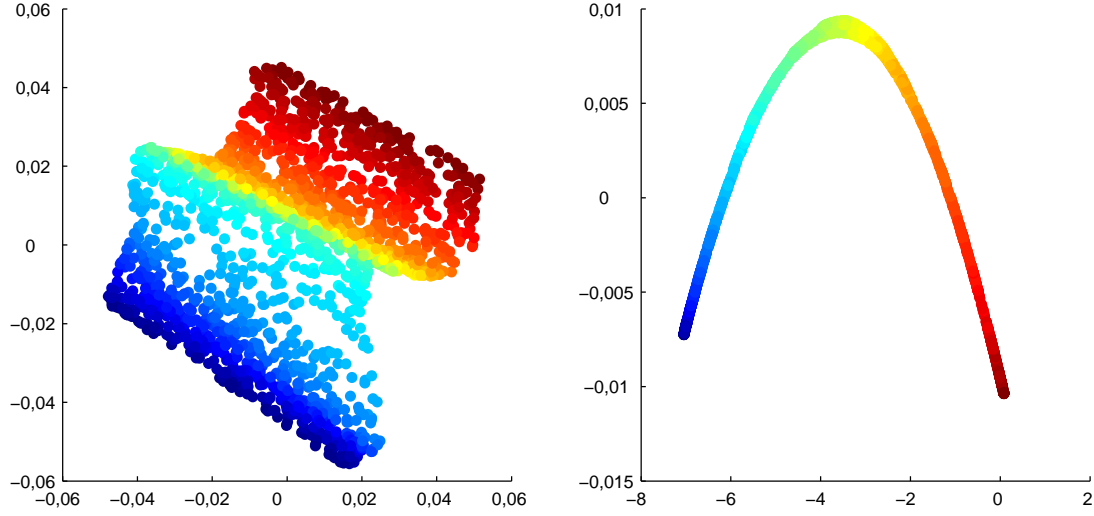


Figura 2.8: Proyección del *Swiss-Roll* con LLE y LE.

A la hora de hacer la proyección de los datos al espacio de baja dimensión, el LLE asume que los datos tienen una relación geométrica que puede ser preservada mediante traslaciones, rotaciones y escalados al realizar la reducción de la dimensión. Para ello, las coordenadas de la proyección deben minimizar el coste de la siguiente función:

$$\Phi(\mathbf{Y}) = \sum_i \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|^2. \quad (2.11)$$

En la figura 2.8 se puede observar el resultado de aplicar este método al *manifold* de referencia *Swiss-Roll*.

Una de las mayores ventajas de este algoritmo es que es simple. Aunque el LLE depende de información local sobre el *manifold*, es capaz de utilizar la información de interconexión entre todos los puntos, lo que garantiza una convergencia óptima del sistema.

Algoritmo LLE

1. Calcular para cada dato de entrada \mathbf{x}_i los k vecinos más próximos.
2. Determinar los pesos \mathbf{W} que minimizan la reconstrucción de los datos a partir de la combinación lineal de los vecinos, mediante la minimización de la ecuación 2.10.
3. Calcular los autovalores de la matriz $(\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$ que se obtiene de resolver la ecuación 2.11. Las coordenadas estimadas vendrán dadas por los autovectores asociados a dicha matriz.

2.3.2. Laplacian Eigenmap

El *Laplacian Eigenmap* (LE) (Belkin y Niyogi, 2001) es un método similar al LLE, al considerar solamente los k vecinos más cercanos, pero considera el problema de vecindad de otra manera al utilizar operadores laplacianos para construir el grafo. El LE está basado en la minimización de la distancia entre los puntos vecinos, estando esta minimización restringida con el fin de evitar la solución trivial. Esta técnica preserva localmente la topología de los datos y tiene una conexión natural al agrupamiento o *clustering*. En la figura 2.8 se puede ver como el LE tiende a agrupar la proyección de los datos.

Este método se basa en una sencilla hipótesis: las relaciones de vecindad que se dan en los datos de entrada \mathbf{X} se mantienen en los datos de salida \mathbf{Y} . Para lograr esto, se construye la matriz binaria de adyacencia \mathbf{A} del espacio de entrada, donde las entradas $a_{ij} \in \{0, 1\}$ indican si los puntos \mathbf{x}_i y \mathbf{x}_j son vecinos, y se define el siguiente criterio:

$$E_{LE} = \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{y}_i - \mathbf{y}_j\|^2 w_{ij}, \quad (2.12)$$

donde los valores w_{ij} de la matriz simétrica \mathbf{W} están relacionados con la matriz de adyacencia del espacio de entrada de la siguiente manera: $w_{ij} = 0$ si $a_{ij} = 0$; si no, $w_{ij} \geq 0$. Para calcular el valor para las entradas distintas de cero se recomienda usar una distribución gaussiana:

$$w_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2T^2}\right). \quad (2.13)$$

El parámetro T puede pensarse como la temperatura de un núcleo de calor en una ecuación de difusión, o de forma más simple, puede considerarse $T = \infty$ si $a_{ij} = 1$.

Algoritmo LE

1. Determinar bien los k vecinos o utilizar un radio de vecindad ε .
2. Construir el correspondiente grafo y la matriz de adyacencia \mathbf{A} .
3. Construir la matriz \mathbf{W} de acuerdo a la ecuación 2.13 y sumar sus columnas para construir la matriz diagonal \mathbf{D} .
4. Obtener la matriz \mathbf{L} , que es el Laplaciano de la matriz \mathbf{W} : $\mathbf{L} = \mathbf{W} - \mathbf{D}$.
5. Normalizar la matriz Laplaciana $\mathbf{L}' = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2}$.
6. Obtener los autovalores y autovectores de la matriz \mathbf{L}' .
7. Obtener la proyección de baja dimensión desnormalizando los autovectores.

Una de las mayores ventajas de este método es la poca cantidad de parámetros que requiere para realizar la proyección, ya que si se considera $T = \infty$, solo es necesario determinar el grado de vecindad. Sin embargo, las proyecciones obtenidas por este método suelen ser muy pobres y suele ser más útil para *clustering* que para reducción de la dimensión.

2.3.3. *Stochastic Neighbor Embedding*

El *Stochastic Neighbor Embedding* (SNE) (Hinton y Roweis, 2002) es un método probabilístico para la proyección de datos, o bien de sus disimilitudes. No trata de preservar distancias entre puntos, como los métodos de escalado multidimensional, sino las probabilidades de que esos puntos sean vecinos. Para ello, se calculan probabilidades condicionales, p_{ij} , que es la probabilidad de que \mathbf{x}_i escoja a \mathbf{x}_j como su vecino mediante una función de densidad gaussiana centrada en el punto \mathbf{x}_i . Esta probabilidad será alta para puntos cercanos, mientras que será casi infinitesimal para puntos lejanos. Matemáticamente, la probabilidad condicional viene dada por:

$$p_{ij} = \frac{\exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_k\|^2}{2\sigma_i^2}\right)}, \quad (2.14)$$

donde σ_i^2 es la varianza de la función gaussiana que está centrada en el punto \mathbf{x}_i . Para los datos proyectados en el espacio de baja dimensión, \mathbf{y}_i e \mathbf{y}_j , es posible calcular una condición de probabilidad similar que se llamará q_{ij} . Por tanto, se modela la similitud de los datos en el espacio de baja dimensión como:

$$q_{ij} = \frac{\exp\left(-\|\mathbf{y}_i - \mathbf{y}_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|\mathbf{y}_i - \mathbf{y}_k\|^2\right)}. \quad (2.15)$$

Si los datos han sido bien proyectados, entonces las probabilidades q_{ij} y p_{ij} deben ser iguales. Por lo tanto, el objetivo del SNE es buscar una proyección de los datos que minimice las diferencias entre estas probabilidades. Para ello, se minimiza la suma de las divergencias de Kullback-Leibler¹ entre todos los puntos:

$$E_{SNE} = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}. \quad (2.16)$$

Dado que la divergencia de Kullback-Leibler no es simétrica (Johnson y Sinanovic, 2001), los errores en las distancias mutuas en el espacio de baja dimensión no son ponderados de igual manera. Esto provoca que la función de coste del SNE se centre en conservar la estructura local de los datos.

El último parámetro a calcular en el SNE es la varianza de la función gaussiana que se centra en cada uno de los puntos del espacio de alta dimensión. Dado que cada punto requeriría un valor de σ_i^2 distinto, el SNE realiza una búsqueda para el valor de la varianza que produce una distribución de probabilidad, P_i , con una perplejidad fija que especifica el usuario. La perplejidad se define como:

$$\text{Perp}(P_i) = 2^{-\sum_j p_{ij} \log_2 p_{ij}}. \quad (2.17)$$

¹Medida de diferencia no conmutativa y no simétrica entre dos distribuciones de probabilidad.

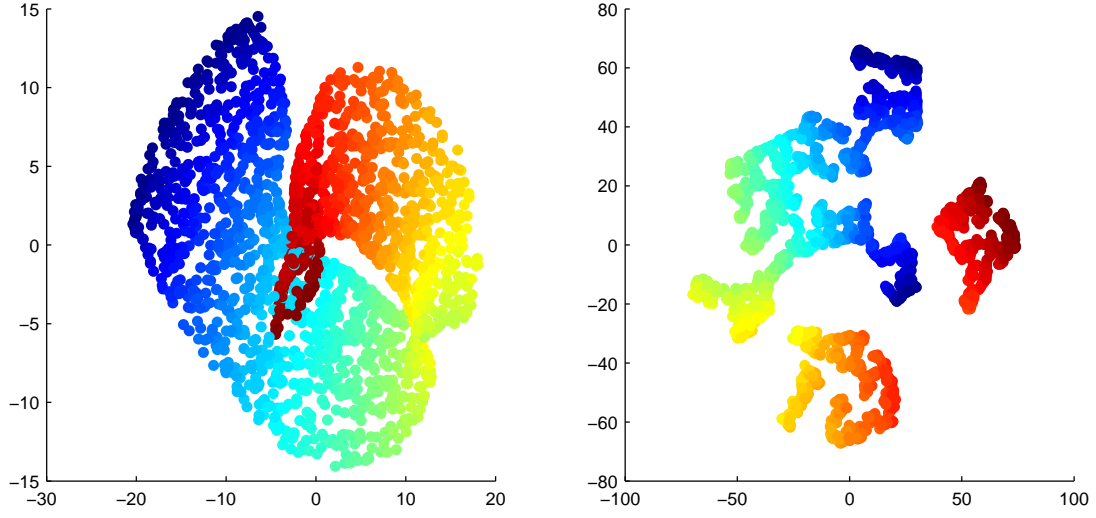


Figura 2.9: Proyección del *Swiss-Roll* con SNE y t-SNE.

La perplejidad puede ser interpretada como una medida de suavizado del número efectivo de vecinos. La eficacia del SNE depende mucho de este parámetro.

La minimización de la función de coste se realiza usando un método de disminución de gradiente. Físicamente, se puede interpretar como un conjunto de muelles entre los puntos del espacio de baja dimensión, de tal manera que la fuerza que estos ejercen depende de las probabilidades de vecindad en el espacio de entrada. El gradiente se inicializa de manera aleatoria y en cada iteración se actualiza conforme a la siguiente ecuación

$$\mathbf{y}_i(t+1) = \mathbf{y}_i(t) - \eta(t) \nabla E_{SNE}(t), \quad (2.18)$$

donde η es la tasa de aprendizaje y $\nabla E_{SNE}(t)$ es la función de minimización de coste que es la derivada parcial de la función de coste respecto a los puntos proyectados $\nabla E_{SNE} = \partial E_{SNE} / \partial \mathbf{y}_i$.

Algoritmo SNE

1. Seleccionar los k vecinos más próximos o calcular la perplejidad de acuerdo a la ecuación 2.17.
2. Calcular las probabilidades p_{ij} del espacio de entrada (ecuación 2.14).
3. Calcular las probabilidades q_{ij} del espacio de salida (ecuación 2.15).
4. Minimizar la función de coste indicada en la ecuación 2.16
5. Actualizar los datos en el espacio de baja dimensión (ecuación 2.18).
6. Repetir desde el paso 3 el número de iteraciones determinado.

Existe una mejora del SNE, denominada ***t-Distributed Stochastic Neighbor Embedding (t-SNE)*** (Van der Maaten y Hinton, 2008), que mejora el algoritmo de SNE en dos puntos:

- Utiliza una versión simétrica de la función de coste (Cook *et al.*, 2007). Por ello, y a diferencia de lo que ocurría en el SNE, las probabilidades son $p_{ij} = p_{ji}$ y $q_{ij} = q_{ji}$ lo cual se logra haciendo $p_{ij} = \frac{p_{ij} + p_{ji}}{2}$.
- Se utiliza una función t-Student en vez de una gaussiana para calcular la similitud entre dos puntos en el espacio de baja dimensión. Esto elimina el problema de amontonamiento de los puntos que se puede dar en el SNE al proyectar cuando existen muchos puntos cercanos en el espacio de alta dimensión.

Una de las mayores ventajas de este métodos es que el t-SNE tiende a conservar más los grupos locales con respecto a otros métodos a la hora de proyectar, al determinar mejor las vecindades durante la ejecución del algoritmo. En la figura 2.9 se puede observar la diferencia de agrupamiento que se produce entre el SNE y el t-SNE al aplicarlo sobre el *Swiss-Roll*. Por el contrario, este método no funciona bien cuando la dimensión de proyección es superior a 3, por lo que solo suele ser útil para visualización. Además, si la dimensión intrínseca de los datos a proyectar es elevada, tampoco obtiene buenos resultados.

2.3.4. Mapas autoorganizados

El ***Self-Organizing Map (SOM)*** (Kohonen, 1982; Kohonen *et al.*, 1996) es una red neuronal no supervisada y autoorganizada, basada en un proceso de aprendizaje competitivo y cooperativo. Realiza una proyección no lineal, ordenada y suave de un espacio de entrada multidimensional continuo sobre un espacio de salida discreto de baja dimensión visualizable. La transformación se hace de forma adaptativa y de un modo topológicamente ordenado (Haykin, 1994).

Cada neurona se representa por un vector p -dimensional de pesos, $\mathbf{w}_i = [w_1, \dots, w_p]$, que se puede interpretar como un vector de coordenadas en el espacio de entrada, y su posición \mathbf{y}_i en la malla de baja dimensión d del espacio de salida. Los vectores prototipo cuantifican los vectores de entrada \mathbf{x} , dividiendo el espacio en una colección finita de regiones de Voronoi, mientras que las neuronas se conectan a las adyacentes por una relación de cercanía que dicta la topología o estructura del mapa.

Durante el entrenamiento, la retícula del SOM se repliega hacia los conjuntos de datos de entrada de tal forma que, como consecuencia del entrenamiento, las coordenadas p -dimensionales de las neuronas de la red se modifican en las zonas donde hay una densidad mayor de datos y por tanto, se localiza una cantidad mayor de neuronas. Las neuronas interaccionan lateralmente con sus adyacentes por medio de la función de vecindad h_{ci} .

El SOM se puede interpretar como un mapa que preserva la topología de los datos del espacio de entrada en una malla. Las coordenadas de las neuronas del SOM en el mapa reticular de baja dimensión lo convierten en un algoritmo de proyección de los datos del espacio de entrada, además, aproxima la función de densidad de los datos de entrenamiento

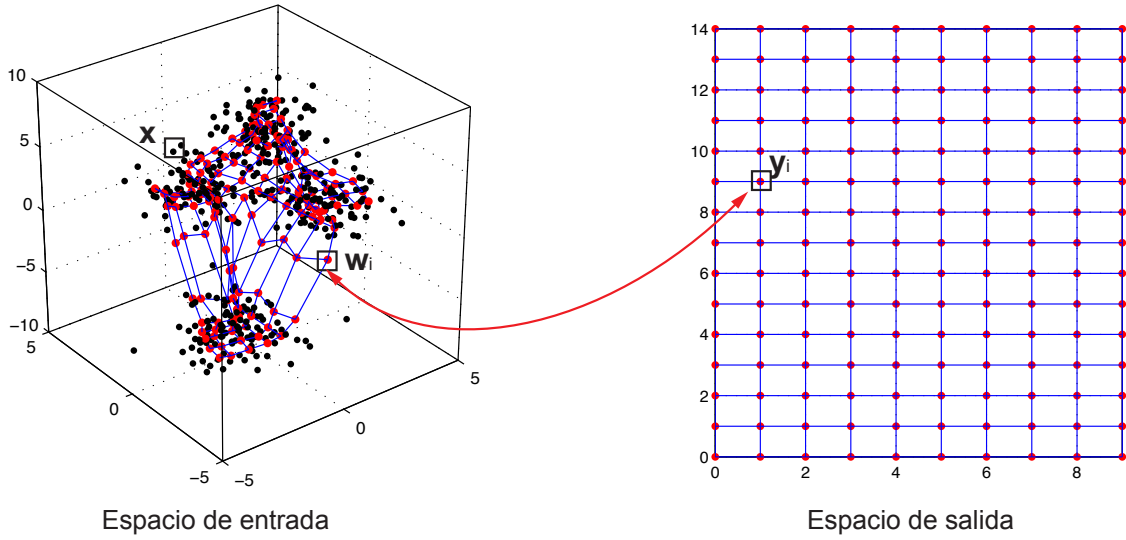


Figura 2.10: Proyección del mapa autoorganizado (SOM).

(las regiones inducidas en el espacio de visualización son tanto mayores cuanto más intenso es el estímulo que representan).

Sea un conjunto de datos en el espacio de entrada \mathbb{R}^p . Cada elemento \mathbf{x} consta de p variables $\mathbf{x} = [x_1, x_2, \dots, x_p]$. El SOM queda entonces definido por una serie de W vectores $\{\mathbf{w}_i\}_{i=1 \dots W}$ que están proyectados en el espacio de salida \mathbb{R}^d . Esta proyección se define como aquel punto de la retícula, \mathbf{y}_i , cuyo homólogo \mathbf{w}_i resulta ser el elemento más próximo a \mathbf{x} en el espacio de entrada de acuerdo con alguna medida de distancia $d(\cdot, \cdot)$ definida en \mathbb{R}^p :

$$S(\mathbf{x}) = \mathbf{y}_c \quad \text{donde} \quad c = \arg \min_i \{d(\mathbf{x}, \mathbf{w}_i)\} . \quad (2.19)$$

Para que este algoritmo de proyección tenga utilidad, es necesario definir los vectores de codificación o prototipos \mathbf{w}_i de manera que la proyección describa adecuadamente la estructura de los datos del espacio de entrada.

Algoritmos de aprendizaje

Como proceso previo al entrenamiento del SOM, los pesos de las neuronas han de inicializarse. Para ello, se pueden utilizar diferentes métodos:

- **Inicialización aleatoria.** Los vectores de pesos de las neuronas toman inicialmente valores diferentes, pequeños y totalmente aleatorios (Haykin, 1994).
- **Inicialización con muestras iniciales.** Los vectores de pesos toman valores correspondientes a alguna de las muestras de datos, por lo que se encuentran en la misma zona del espacio que los datos de entrada (Hollmén, 1996).
- **Inicialización lineal.** Los vectores de pesos se inicializan de forma ordenada y lineal, a lo largo de los mayores autovectores (Kohonen, 2001).

Una vez inicializado, se procede a aplicar el método de aprendizaje. Inicialmente se utilizaba un método conocido como *Vector Quantization* (VQ), que busca determinar los vectores de las neuronas de manera que su distribución de probabilidad se parezca lo más posible a la de los datos de entrada. Para ello, minimiza el **error de cuantización promediado** (Kohonen, 1990a):

$$E_{VQ} = \int f[d(\mathbf{x}, \mathbf{w}_c)] p(\mathbf{x}) d\mathbf{x}, \quad (2.20)$$

donde f es una función monótona creciente de la distancia $d(\cdot, \cdot)$. Esta función de coste evalúa la distancia promediada entre los vectores del espacio de entrada \mathbf{x} y los vectores de codificación que los representan.

La minimización de la función de coste 2.20 no produce mapas ordenados. La solución a esta función padece una importante indeterminación, al ser igualmente válida cualquier permutación entre los vectores de codificación hallados. Kohonen (1990b, 2001) propone la introducción de un núcleo de suavizado local, h_{ci} , función de la distancia en el espacio de salida entre las unidades \mathbf{y}_c e \mathbf{y}_i :

$$E_{KOH} = \int h_{ci} \cdot f[d(\mathbf{x}, \mathbf{w}_c)] p(\mathbf{x}) d\mathbf{x}. \quad (2.21)$$

La introducción del núcleo h_{ci} induce un orden en los vectores de codificación obtenidos al minimizar la función de coste. El carácter competitivo del algoritmo anterior se combina con un carácter cooperativo: no solo debe mejorar la unidad ganadora, sino también sus vecinas. Esta nueva regla de aprendizaje arrastra a la neurona ganadora y a sus neuronas vecinas hacia cada nuevo vector de entrada, como una red flexible que se pliega sobre el conjunto de datos de entrada (ver figura 2.11).

Existen varios algoritmos que permiten obtener una solución aproximada para las \mathbf{w}_i . El algoritmo del SOM, propuesto por Kohonen, está compuesto por dos etapas. La primera etapa, llamada etapa de competición, consiste en buscar la neurona ganadora, *Best Matching Unit* (BMU), por medio de un proceso competitivo por el que se selecciona aquella más cercana:

$$c(t) = \arg \min_i \|\mathbf{x}(t) - \mathbf{w}_i(t)\|, \quad i = 1, 2, \dots, W \quad (2.22)$$

donde W es el número de neuronas y t indica el tiempo. Normalmente, se utiliza la norma euclídea para calcular la distancia. La segunda etapa, llamada etapa de cooperación, está constituida por un proceso cooperativo y adaptativo, mediante el cual, los pesos de las neuronas ganadoras y sus vecinas se adaptan en función del nuevo vector de entrada.

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t)h_{ci}(t)[\mathbf{x}(t) - \mathbf{w}_i(t)]. \quad (2.23)$$

El núcleo h_{ci} es una función decreciente con la distancia definida en la retícula del espacio de salida. Esta función es generalmente gaussiana (ecuación 2.24) pero se pueden utilizar otros tipos de funciones como es la función burbuja, la función *cut-Gauss* o la función de Epanechnikov (Vesanto y Alhoniemi, 2000).

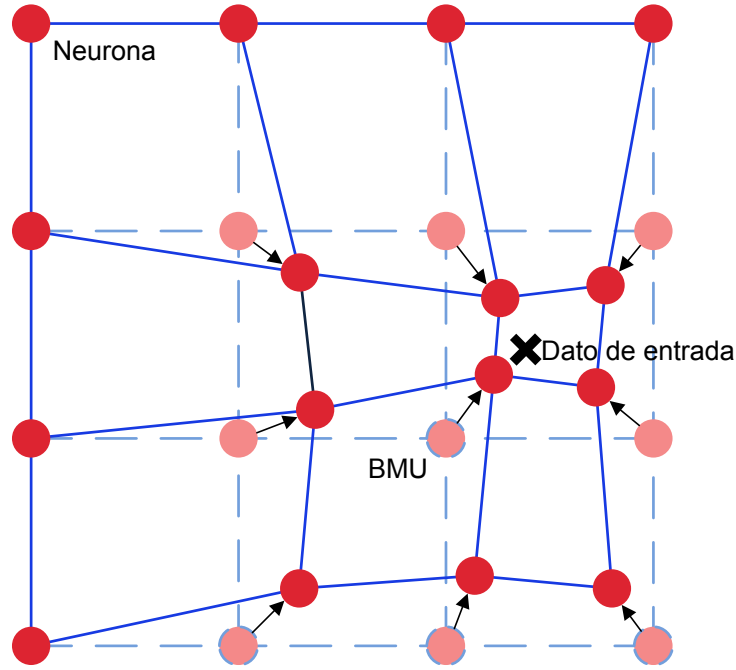


Figura 2.11: Esquema del proceso de adaptación en el mapa autoorganizado (SOM).

$$h_{ci}(t) = \exp \left(-\frac{\|\mathbf{y}_i - \mathbf{y}_c\|^2}{2\sigma(t)^2} \right) \quad (2.24)$$

El parámetro $\sigma(t)$ controla la cobertura efectiva de la función de vecindad. La relación de vecindad dicta la topología del mapa, que generalmente es rectangular o hexagonal. Este parámetro, junto con la velocidad de aprendizaje $\alpha(t)$, rige la adaptación y es crítico para el éxito o fracaso de la proyección y, en general, su valor va disminuyendo con el tiempo (*annealing*). Cualquier función de vecindad debe satisfacer los siguientes requisitos:

- Su mayor valor se encuentra en la neurona ganadora c , para la que la distancia $\|\mathbf{y}_i - \mathbf{y}_c\|$ es cero.
- Decece a cero cuando la distancia tiende a infinito.

Existe una versión de entrenamiento del SOM por lotes o *batch* en el que se entrena con todos los datos simultáneamente. La adaptación se realiza mediante una media ponderada de los vectores de entrada de forma que el nuevo peso de cada vector es la función de vecindad $h_{c_j i}(t)$ correspondiente a su neurona ganadora c_j . Para ello, se divide el conjunto de datos de entrada según las regiones de Voronoi establecidas por los vectores prototipo, es decir, de manera que cada vector de entrada pertenezca al conjunto de los que tienen la misma neurona más próxima.

El algoritmo *batch* es de los más utilizados, dada su alta eficiencia computacional en comparación con otros métodos.

Algoritmo *batch*

1. Obtener los índices de las unidades vencedoras para cada dato de ejemplo \mathbf{x}_j .
Etapa de competición.

$$c_j = \arg \min_i \|\mathbf{x}_j - \mathbf{w}_i\|, \quad i = 1, 2, \dots, W, \quad j = 1, 2, \dots, N$$

2. Obtener los nuevos \mathbf{w}_i mediante la media ponderada. Etapa de cooperación.

$$\mathbf{w}_i(t+1) = \frac{\sum_{j=1}^N h_{c_j i}(t) \mathbf{x}_j}{\sum_{j=1}^N h_{c_j i}(t)}$$

3. Repetir los pasos 1 y 2, decreciendo suavemente con el tiempo la amplitud del campo receptivo (σ) de h_{ci} hasta la convergencia.

Propiedades del SOM

El mapa autoorganizado presenta una serie de propiedades que lo hacen idóneo para la visualización y exploración de grandes volúmenes de datos.

Cuantización de vectores. El SOM divide el espacio de entrada en una colección finita de regiones de Voronoi. De hecho, sus ecuaciones presentan cierto parecido con las de los algoritmos Linde-Buzo-Gray (LBG) y *k-means*, los cuales son utilizados para el *clustering* de datos. Sin embargo, debido a la incorporación del concepto de vecindad, el SOM no consigue una transferencia óptima de información ya que la aproximación de los vectores prototipo a los datos está sujeta a una restricción. Esto no constituye una gran desventaja ya que hace posible preservar la topología y provoca que la retícula de salida sea una representación de dimensión reducida de los datos de entrada.

Aproximación de la función de densidad de los datos. El SOM dedica más neuronas a aquellas regiones en las que la función de densidad de probabilidad de los datos es mayor. Por lo tanto, regiones con más densidad de probabilidad del espacio de entrada se proyectarán sobre zonas más amplias. En definitiva, el área dedicada a la representación de una característica determinada es, en cierta medida, proporcional a la frecuencia estadística de aparición de dicha característica. El factor de magnificación describe la relación entre la densidad de los datos y la densidad de vectores prototipo en la salida. Sin embargo, el SOM tiende a realizar una representación mayor de regiones de baja densidad y una menor de regiones de alta densidad (Haykin, 1994).

Ordenación topológica. Los vectores del SOM se distribuyen a lo largo del espacio de entrada de manera que, además de distribuirse de forma similar a como lo hacen los datos, lo hacen de forma ordenada, en el sentido de que las relaciones de vecindad definidas en el espacio de salida prevalecen en los vectores de codificación del espacio de entrada tras la

convergencia. Por tanto esta ordenación indica similitudes en la estructura de vecindad de los conjuntos de datos de entrada y de salida. Esta propiedad es una consecuencia directa del núcleo h_{ci} , que es un factor que implica el movimiento de adaptación de no solo la neurona vencedora sino también de sus vecinas. Esto hace que el SOM realice una proyección local correcta, pero no optimiza la proyección de forma global como otros métodos (Kaski, 1997).

Reducción de la dimensionalidad. Es una propiedad consecuencia de las anteriores. Dado que el SOM es capaz de generar una salida de baja dimensión en una estructura de malla fija que conserva la topología y es capaz de aproximar la geometría de la función de densidad de los datos, el mapa resultante puede indicar en las posiciones de la malla 2D todas las características del espacio de entrada. El SOM realiza una proyección sobre una superficie que se deforma ajustándose a los datos. Se trata en cierto sentido de un análisis de componentes principales no lineal.

Capacidad de visualización. El SOM proporciona una visualización compacta y ordenada de la información, debido a que los vectores se proyectan en un espacio delimitado y fijo en la malla de salida (Kaski, 1997). Si se combina con otros métodos de visualización, como por ejemplo con un método de escalado multidimensional, se puede obtener información de los datos de forma sencilla y rápida. Las cualidades del SOM hacen que proporcione fiabilidad (*trustworthiness*) en la visualización de los datos de entrada (Venna, 2007).

Uno de los principales problemas del SOM radica en la ausencia de una función de energía. Esto impide que se pueda realizar una minimización por descenso de gradiente para optimizar el algoritmo (Erwin *et al.*, 1992). Otro de los problemas reside en que el cálculo de la BMU no se tiene en cuenta en la función de error, lo que impide obtener una función de energía válida. Por otra parte, no se puede garantizar la convergencia del algoritmo cuando la dimensión de la malla de salida es mayor de uno.

Herramientas de visualización

A diferencia de las técnicas previas de reducción de la dimensionalidad, el espacio reticular topológicamente ordenado del SOM genera visualizaciones en las cuales las neuronas tienen las mismas distancias unas de otras. Sin embargo, con las visualizaciones del SOM se pueden mostrar propiedades escalares o variables del espacio de entrada, distancias entre neuronas, etc. Las coordenadas de las neuronas permiten determinar la posición y el uso de un código de color permite mostrar de forma intuitiva y homogénea los valores.

Debido a estas cualidades, se han propuesto a lo largo de la bibliografía una serie de visualizaciones que trabajan tanto con la información de las neuronas como con sus coordenadas para mostrar distintas propiedades y que no son directamente aplicables al resto de técnicas. A continuación se muestran algunas de las técnicas de visualización más destacadas que tienen como base el SOM.

Plano de componentes. También conocidos como mapa de características, permite describir la variable j -ésima del proceso, construyendo el mapa de la componente j -ésima de las neuronas \mathbf{w}_i (Tryba *et al.*, 1989). Cada plano representa el valor de una variable del vector

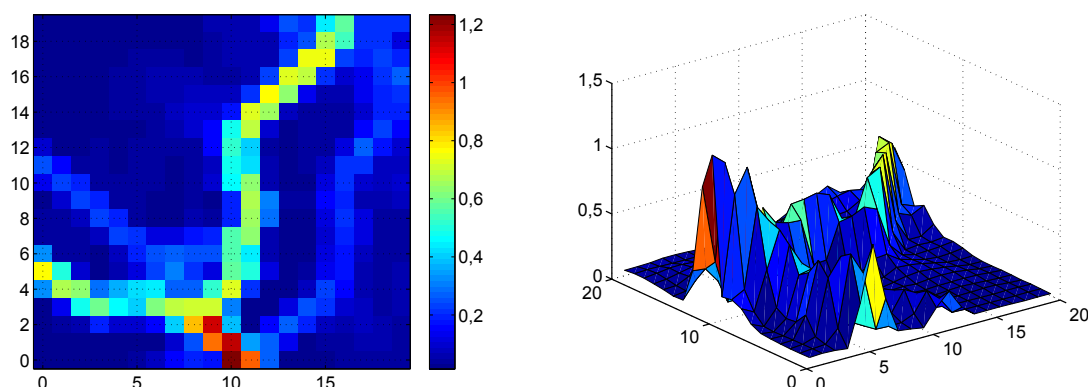


Figura 2.12: Matriz de distancias 2D y 3D.

prototipo en cada nodo del SOM. Es habitual utilizar un pseudo-color para representar el valor asignado a cada neurona. Los planos de componentes juegan un papel importante en la detección de correlaciones e incluso variables correlacionadas parcialmente se pueden detectar con una simple inspección visual.

Matriz de distancias unificada. También conocida como *u-matrix* o **mapa de distancias** (Ultsch y Siemon, 1990), es un mapa que permite describir la estructura de *clusters* de los datos, representando como propiedad escalar, por medio de colores, la distancia promedio de cada neurona a sus vecinas más próximas, de acuerdo con una función de vecindad (ver figura 2.12). La distancia promedio guarda una estrecha relación con la densidad de las neuronas, de tal manera que una región del espacio de salida con pequeñas distancias interneuronales representa zonas densamente pobladas en el espacio de entrada, que se corresponderían con un *cluster*. Por otra parte, zonas con baja densidad pueden verse como zonas de separación entre los *clusters*. La inspección visual del mapa a menudo permite determinar de forma manual las agrupaciones de datos, por lo que se han propuesto métodos automáticos de etiquetado (Domínguez, 2003).

Histogramas. La BMU de un dato del proceso representa el estado al que pertenece ese dato según el modelo (Vesanto, 1999). Para un conjunto de datos, los múltiples vectores BMU permiten generar un histograma de dichos datos. Comparando distintos histogramas se pueden analizar las similitudes o diferencias entre conjuntos de datos.

Mapa de estados. Es un mapa en el que los colores o etiquetas de las neuronas dependen del *cluster* al que pertenezcan (Fuertes *et al.*, 2005). En el caso de que se analice un proceso, cada grupo puede considerarse como un estado del mismo (Ahola *et al.*, 1999). Existe una modificación denominado mapa de etiquetas en el que se identifica cada estado con una etiqueta (Fuertes, 2006).

Proyección de la trayectoria. La proyección del dato actual en el SOM mediante el cálculo de la BMU permite identificar el estado del proceso. Si se conecta la secuencia ordenada de la proyección de estos puntos, se genera una trayectoria sobre el mapa de visualización (ver figura 2.13) que da idea de la evolución del proceso (Kasslin *et al.*, 1992; Tryba y Goser, 1991). A partir de esta trayectoria es posible determinar qué *clusters* o

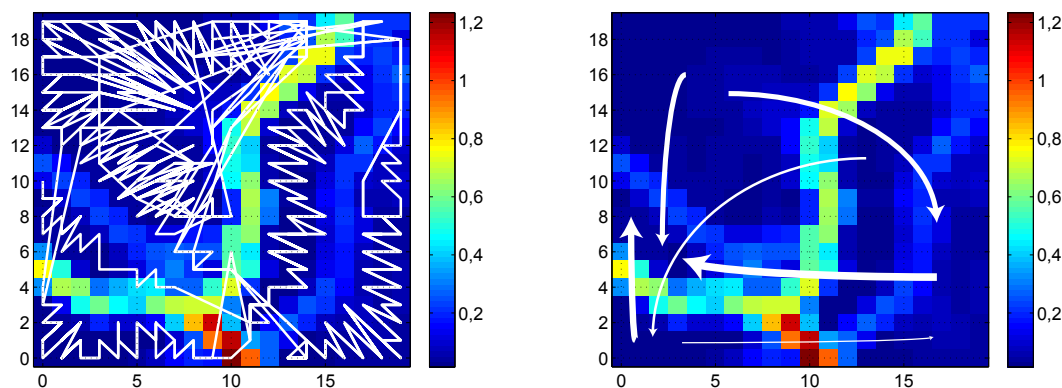


Figura 2.13: Mapa de trayectorias y de transiciones.

condiciones de proceso son accesibles desde una dada y cuál es la probabilidad de transición entre ellas. Además los vectores de dicha trayectoria permiten construir modelos dinámicos de evolución del sistema (Fuertes *et al.*, 2007).

Mapa dinámico de transiciones. Partiendo del mapa de estados y de las trayectorias en el mapa, se pueden calcular probabilidades de transición entre condiciones de proceso. Estas probabilidades se pueden representar mediante flechas con grosores proporcionales a dicha probabilidad (Fuertes *et al.*, 2007). La visualización de este mapa junto con el estado actual del sistema permite no solo la detección de fallos sino la predicción de futuros estados del proceso. En la figura 2.13 se muestra un mapa dinámico de transiciones.

Representación de los residuos. El concepto de residuo se puede relacionar con el error de cuantización (Kohonen *et al.*, 1996) entre un vector de entrada y el vector prototipo correspondiente a la BMU de un SOM entrenado. Este residuo da información de las desviaciones o fallos en el proceso y su representación gráfica en el tiempo es útil cuando la dimensión del espacio de entrada es alta (Díaz y Hollmén, 2002).

Modelo dinámico de trayectorias. Para este propósito, se utiliza una variante del SOM llamado *Kernel Regression SOM* (KR-SOM), que consiste en una modificación del algoritmo SOM en la que la secuencia de activaciones de la BMU forma una trayectoria 2D continua más fácil de analizar. Esta trayectoria, a su vez, se utiliza como entrada para otro SOM (Fuertes *et al.*, 2007). El aprendizaje de los vectores de flujo de la trayectoria permite definir herramientas de visualización más sensibles a cambios leves (Fuertes *et al.*, 2007) que, por ejemplo, permiten representar las discrepancias entre la trayectoria modelada por el SOM y la generada por otra ejecución del sistema. Así, se pueden detectar desviaciones en la dirección o cambios de velocidad o aceleración que indiquen un comportamiento anómalo.

Mapa de diferencias. Estos mapas permiten la comparación de procesos que se rigen por el mismo patrón de funcionamiento para conocer los puntos con un distinto comportamiento entre los procesos (Fuertes, 2006). Para ello se parte de dos instancias entrenadas del SOM y se representa una función diferencia que asigna a cada neurona del modelo B la diferencia entre su vector prototipo y el de la BMU de esta neurona en el

modelo A . En estos mapas se puede incorporar información sobre las transiciones, lo que da lugar a los mapas de diferencias dinámicos.

2.4. Cuantización y *clustering* de vectores

La cuantización de vectores tiene como objetivo realizar una compresión de datos en aquellas aplicaciones que toleran cierta distorsión. La idea fundamental detrás de la cuantización es aproximar los vectores de entrada utilizando un número pequeño de vectores prototipo (Cherkassky y Mulier, 2007). Una cuantización Q de vectores de dimensión p es una transformación del espacio euclídeo p -dimensional \mathbb{R}^p en un conjunto finito K de **vectores prototipo**, que recibe el nombre de *codebook*, $\{\mathbf{q}_i\}_{i=1\dots K}$. Habitualmente, se utiliza un K mucho menor que el número de datos del conjunto de entrada. Generalmente se trabaja con el error cuadrático medio o norma ℓ^2 como medida de la distorsión que ha de minimizarse. Los vectores prototipo parten \mathbb{R}^p en regiones, que se definen como

$$R_i = \{\mathbf{x} \in \mathbb{R}^p : Q(\mathbf{x}) = \mathbf{q}_i\}. \quad (2.25)$$

Un vector reproduce de forma óptima una región R_i cuando es el centroide

$$\mathbf{q}_i = \frac{\sum_{\mathbf{x}_j \in R_i} \mathbf{x}_j}{|R_i|}, \quad (2.26)$$

donde $|\cdot|$ denota la cardinalidad (número de elementos) del conjunto. La partición es de Voronoi cuando para un conjunto fijo de vectores de reproducción, la partición óptima se construye de la siguiente manera:

$$Q(\mathbf{x}) = \mathbf{q}_i \Leftrightarrow \|\mathbf{x} - \mathbf{q}_i\| \leq \|\mathbf{x} - \mathbf{q}_r\|, \forall r \neq i. \quad (2.27)$$

En cuanto al *clustering* o agrupamiento, es similar a la cuantización, diferenciándoles únicamente el objetivo, pues este busca agrupaciones interesantes de los datos de entrada (Xu y Wunsch, 2005). El resultado es directamente una serie de particiones Q de los datos del espacio de entrada, pudiendo identificarse cada región por un vector que representa el centroide del *cluster* \mathbf{q} . Trabajar con los centroides obtenidos por una técnica de *clustering* es similar a hacerlo con los vectores prototipo de una cuantización. El objetivo básico de la cuantización es la representación de los datos mientras que el del *clustering* es su interpretación. La mayoría de los algoritmos pueden servir tanto para un propósito como para el otro.

Los métodos más utilizados son el *k-means*, el *Neural Gas* y los algoritmos jerárquicos. Dado que los métodos de *clustering* requieren como parámetro el número de *clusters* en los que se van a partir los datos, es necesario tener un indicador que permita determinar lo buena que es esa partición (Milligan y Cooper, 1985). Para comparar la eficacia del agrupamiento en función del número de *clusters* seleccionados para su partición, existen una serie de índices que permiten comparar los resultados entre sí y seleccionar el número óptimo. Dentro de los índices existentes los más utilizados son:

- **Índice Davies-Bouldin.** Este índice (Davies y Bouldin, 1979) es una función del coeficiente entre la separación de los puntos dentro del mismo *cluster* y la separación entre los *clusters*:

$$DB_K = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{\text{diam}(Q_i) + \text{diam}(Q_j)}{d(Q_i, Q_j)} \right), \quad (2.28)$$

donde $\text{diam}(Q_i)$ es la función de intracuster o diámetro interno del *cluster* que recoge todas las distancias entre los puntos pertenecientes a un mismo grupo y que se define como

$$\text{diam}(Q_i) = \sqrt{\frac{1}{N_i} \sum_{\mathbf{x} \in Q_i} \|\mathbf{x} - \mathbf{q}_i\|^2}, \quad (2.29)$$

siendo N_i el número de puntos y \mathbf{q}_i el centroide *cluster* Q_i . El término $d(Q_i, Q_j)$ se refiere a la distancia entre los centroides.

Este índice toma valores pequeños para el caso de *clusters* compactos y con una elevada distancia entre ellos, por tanto, tiene valores pequeños para los casos en los que se producen buenos agrupamientos.

- **Índice de Dunn.** Este índice (Dunn, 1974) se basa en la idea de diferenciar los *clusters* que son compactos de aquellos en los que los puntos están bien separados:

$$DU_K = \min_{i \neq j} \left(\frac{S(Q_i, Q_j)}{\max_k \text{diam}(Q_k)} \right), \quad (2.30)$$

donde $S(Q_i, Q_j)$ mide la disimilitud entre los *clusters* Q_i y Q_j , la cual viene definida por

$$S(Q_i, Q_j) = \min_{\substack{\mathbf{x} \in Q_i \\ \mathbf{y} \in Q_j}} \|\mathbf{x} - \mathbf{y}\|, \quad (2.31)$$

en cuyo caso el diámetro viene definido por

$$\text{diam}(Q) = \max_{\mathbf{x}, \mathbf{y} \in Q} \|\mathbf{x} - \mathbf{y}\|. \quad (2.32)$$

Este índice toma valores grandes cuando las distancias entre los *clusters* son grandes y las distancias internas de un grupo son pequeñas. Por ello, el número de grupos K escogido será aquel que maximice este índice.

- **Índice Silhouette.** Este índice (Rousseeuw, 1987) calcula la anchura del contorno para cada muestra, la media del contorno para cada *cluster* y el contorno para todo el conjunto de datos. Utilizando este índice, cada grupo puede ser representado por su contorno, el cual se basa en la comparación de su empaquetamiento y separación. La silueta $\text{silh}(i)$ viene dada por la ecuación:

$$\text{silh}(i) = \frac{a(i) - b(i)}{\max(a(i), b(i))}, \quad (2.33)$$

donde $a(i)$ es la distancia media del punto i a todos los puntos del mismo *cluster* y $b(i)$ es el mínimo de las distancias medias entre el punto i y todos los puntos de otros *clusters*.

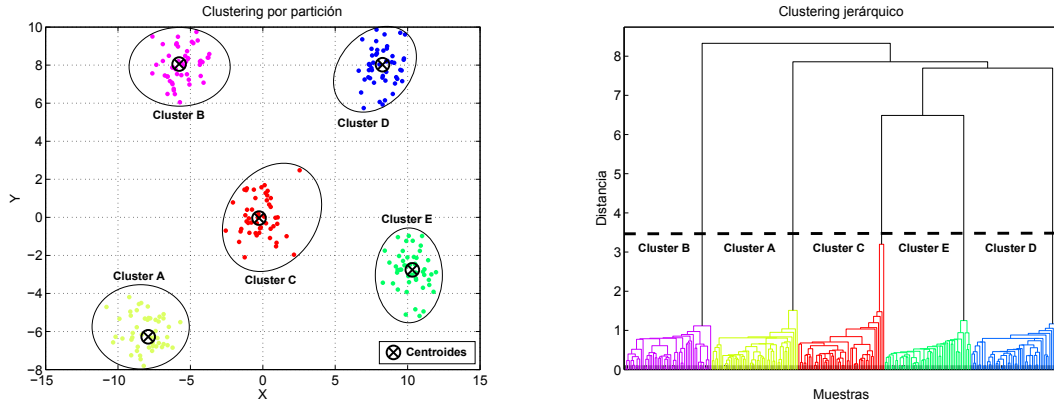


Figura 2.14: Método de agrupamiento por *k-means* y jerárquico.

2.4.1. *k-means*

Es un método clásico de *clustering* basado en el agrupamiento por partición, dividiendo el conjunto de datos de entrada en un número de grupos determinado (MacQueen, 1967). Este algoritmo busca de forma iterativa el número de grupos óptimo que realiza una partición de mínima varianza en el espacio de datos de entrada, minimizando la función de coste E :

$$E_{KM} = \sum_{k=1}^K \sum_{\mathbf{x}_i \in Q_k} \|\mathbf{x}_i - \mathbf{q}_k\|^2, \quad (2.34)$$

donde \mathbf{q}_k es el centroide del *cluster* Q_k .

Algoritmo *k-means*

1. Inicializar todos los centroides \mathbf{q}_k de forma aleatoria o aplicando conocimiento *a priori*.
2. Asignar cada dato de entrada \mathbf{x}_i al grupo Q_l más próximo, es decir, $\mathbf{x}_i \in Q_l$ si $\|\mathbf{x}_i - \mathbf{q}_l\| < \|\mathbf{x}_i - \mathbf{q}_k\|$, $i = 1, 2, \dots, N$, $k, l = 1, 2, \dots, K$, $i \neq l$.
3. Calcular los centroides, teniendo en cuenta la partición actual. Esto es, $\mathbf{q}_k = \frac{1}{|Q_k|} \sum_{\mathbf{x}_i \in Q_k} \mathbf{x}_i$, donde $|Q_k|$ representa el número de datos dentro del grupo k .
4. Repetir los pasos 2 y 3 hasta que los centroides dejen de modificarse definitivamente.

Uno de los inconvenientes de este método es que hay que proporcionar de antemano el número de grupos que se quieren obtener, cuando generalmente es lo que se quiere conocer. Por ello es necesario realizar varias iteraciones y tratar de minimizar el error (Halkidi *et al.*, 2001). Otro de los problemas es que está limitado a la separación de grupos con formas no convexas. A esto hay que añadir el gran coste computacional cuando el número de datos es elevado.

Existe gran cantidad de métodos similares y variantes del *k-means* que se utilizan para *clustering*. Entre ellos destaca el LBG (Linde *et al.*, 1980) que en lugar de buscar las agrupaciones, trata de minimizar la norma ℓ^2 y garantiza alcanzar un mínimo local. El problema de este método es que su solución depende en gran medida de la inicialización del algoritmo, por lo que es necesario ejecutarlo varias veces con distinta inicialización.

2.4.2. *Neural Gas*

El *Neural Gas* (Martinetz *et al.*, 1993) también es un método de agrupamiento por partición. Es una técnica muy similar al SOM ya que ambas preservan la topología, pero sin la imposición de una estructura de vecindad predefinida, lo que permite trabajar mejor en espacios de entrada complicados. El *Neural Gas* permite obtener un número finito K de vectores prototipo preservando la función de densidad de probabilidad del espacio de entrada. La actualización no se aplica en función de la distancia al vector ganador sino en función del orden de distancia de los vecinos $(\mathbf{q}_{i_0}, \mathbf{q}_{i_1}, \dots, \mathbf{q}_{i_{K-1}})$ siendo \mathbf{q}_{i_0} el más cercano al dato \mathbf{x} . La adaptación de los vectores prototipo se calcula en cada paso como:

$$\mathbf{q}_{i_k}(t+1) = \mathbf{q}_{i_k}(t) + \alpha(t)e^{-k/\lambda}[\mathbf{x}(t) - \mathbf{q}_{i_k}(t)], \quad (2.35)$$

donde $\alpha(t)$ es la tasa de aprendizaje, λ es la función de vecindad y el índice k es el orden del vector prototipo \mathbf{q}_i con respecto al dato \mathbf{x} . Si se compara con la ecuación de aprendizaje del SOM (ecuación 2.23) se puede comprobar la similitud con la misma.

Este método consigue mejores resultados de convergencia que el *k-means*; sin embargo, al igual que este algoritmo sufre el problema de que la falta de un espacio de salida limita su uso para visualización.

Algoritmo *Neural Gas*

1. Inicializar los centroides \mathbf{q}_i de forma aleatoria o aplicando conocimiento *a priori*.
2. Calcular el orden de los prototipos con respecto a cada dato de entrada.
3. Aplicar la ecuación 2.35 para cada uno de los datos.
4. Repetir los dos pasos anteriores el número de iteraciones que se haya determinado.

2.4.3. Agrupación jerárquica

Más que una técnica, es una familia de métodos (Gordon, 1987). Ya que no existe un método concreto que destaque sobre todos los demás, es mejor hablar del método en general. Este tipo de agrupación busca construir una estructura de grupos mediante una secuencia de particiones anidadas, a partir de los datos de entrada. Se genera un árbol de grupos donde los datos se organizan siguiendo una estructura jerárquica en función de algún criterio de enlace o vínculo. Los principales criterios son:

- **Single linkage.** La disimilitud entre dos grupos δ_{SL} viene dada por la distancia más pequeña entre todos los datos pertenecientes a ambos grupos.

$$\delta_{SL} = \min \{d(\mathbf{x}_i, \mathbf{x}_j)\}, \mathbf{x}_i \in Q_k, \mathbf{x}_j \in Q_l. \quad (2.36)$$

- **Complete linkage.** La disimilitud entre dos grupos δ_{CL} viene determinada por la distancia más grande entre todos los datos pertenecientes a ambos grupos.

$$\delta_{CL} = \max \{d(\mathbf{x}_i, \mathbf{x}_j)\}, \mathbf{x}_i \in Q_k, \mathbf{x}_j \in Q_l. \quad (2.37)$$

- **Average linkage.** La disimilitud entre dos grupos δ_{AL} se define como la distancia promedio entre todos los pares de datos de ambos grupos.

$$\delta_{AL} = \frac{1}{|Q_k||Q_l|} \sum_{\mathbf{x}_i \in Q_k} \sum_{\mathbf{x}_j \in Q_l} d(\mathbf{x}_i, \mathbf{x}_j). \quad (2.38)$$

Para las tres ecuaciones anteriores, $d(\mathbf{x}_i, \mathbf{x}_j)$ es una medida de distancia entre los datos de entrada \mathbf{x}_i y \mathbf{x}_j . Q_k y Q_l son dos grupos determinados, mientras que $|Q_k|$ y $|Q_l|$ indican el número de datos dentro del grupo k y l , respectivamente.

Algoritmo Jerárquico

1. Constituir tantos grupos como datos de entrada.
2. Calcular las disimilitudes o similitudes entre grupos $d(Q_k, Q_l)$, de acuerdo a alguno de los criterios de las ecuaciones 2.36, 2.37 y 2.38.
3. Fusionar los dos grupos más similares para construir un grupo mayor.
4. Repetir los dos pasos anteriores hasta formar un grupo único.

Normalmente, el sistema de visualización de un método jerárquico es mediante un dendrograma, también conocido como árbol binario. Este tipo de representación se va escindiendo en función del número de grupos en los que se van separando los datos a la hora de determinar los grupos. La distancia entre los grupos viene dado por la altura del mismo. El número de grupos se obtiene “cortando” el dendrograma en un determinado nivel.

El agrupamiento jerárquico es útil para visualizar la estructura potencial de grupos cuando existen relaciones en los datos de entrada. Como inconvenientes, hay que destacar su gran coste computacional y su sensibilidad al ruido. Generalmente se suelen combinar los dos métodos de tal manera que un método jerárquico dé idea del número de grupos a obtener y el método por partición genere vectores prototipos más precisos.

Predicción de series temporales

La predicción utilizando series temporales permite, mediante funciones de aproximación, estimar valores futuros de una secuencia de observaciones de valores pasados. El análisis de series temporales constituye un campo de trabajo muy estudiado a lo largo de la bibliografía y diversas técnicas como pueden ser redes neuronales o descomposiciones de Fourier han sido probadas. Estas técnicas no solo abarcan métodos de generación de modelos sino también técnicas de selección de variables y del tamaño de la ventana de predicción (Gooijer y Hyndman, 2006; Box y Jenkins, 1970).

Dentro de las numerosas técnicas de predicción de series temporales existen métodos basados en *clustering* de vectores que buscan reducir el número de datos utilizados y obtener los más representativos entrenando un modelo para cada uno de los *clusters*, que se supone representativo (Fan y Chen, 2006). Además del *clustering*, también destaca el empleo de técnicas de reducción de la dimensionalidad, entre las que destaca el SOM que ha sido utilizado ampliamente a lo largo de la literatura, bien sea utilizando el algoritmo original o diversas modificaciones del mismo (Príncipe *et al.*, 1992).

En cuanto a otras técnicas de reducción de la dimensionalidad, como son el CCA, Isomap, LLE, etc., han sido utilizadas en menor medida. Los usos que existen en la literatura buscan generar series temporales con los datos en el espacio de salida de menor dimensión y predecir las nuevas variables en este espacio en vez de hacerlo con los datos de entrada. Al reducir el número de variables se reduce la complejidad del problema de predicción (Chen *et al.*, 2008).

Dado que estas técnicas descritas también se usan para el análisis de los datos y la generación de visualizaciones, como se comentó en el capítulo 2, a lo largo de este capítulo se les presta una mayor atención con el fin de mostrar su viabilidad en la predicción de series temporales.

3.1. Series temporales

Una serie temporal es una secuencia ordenada $\{x(t)\}_{t=1}^N$ de N observaciones de una variable escalar $x \in \mathbb{R}$ (Barreto, 2007). La predicción de los valores futuros de una serie temporal pasa por conocer la dinámica de la misma. Sin embargo, dado que esta dinámica suele ser desconocida, el propósito principal en el estudio de series temporales consiste en

realizar una serie de pruebas que permitan obtener el modelo dinámico que mejor se ajuste a la serie. Una vez que se ha conseguido obtener este modelo, es posible aplicar el mismo a nuevos datos con el fin de conocer la futura evolución de la serie temporal. Una hipótesis general al trabajar con series temporales es que estas pueden ser descritas por medio de un modelo regresivo no lineal,

$$\begin{aligned} x(t+1) &= f[x(t), x(t-1), \dots, x(t-(m-1))] + \varepsilon(t) \\ &= f[\mathbf{x}^-(t)] + \varepsilon(t), \end{aligned} \quad (3.1)$$

donde $\mathbf{x}^-(t) \in \mathbb{R}^m$ es el vector regresor de orden m de valores pasados, $f(\cdot)$ es una función no lineal de sus argumentos y ε es una variable aleatoria que se incluye para tener en cuenta variaciones y ruido, el cual se considera generalmente como ruido gaussiano. A la hora de estudiar las series temporales se pretende obtener la estructura del mapeado $f(\cdot)$ a partir del conocimiento de las muestras pasadas de la serie temporal que sí son conocidas. Este modelo permitirá obtener la predicción de las muestras futuras en función de un regresor de entrada conocido,

$$\hat{x}(t+1) = f[\mathbf{x}^-(t) | \Theta], \quad (3.2)$$

donde Θ se refiere al vector de parámetros ajustables del modelo. La función $f[\cdot | \Theta]$ puede ser implementada mediante diferentes técnicas, como pueden ser modelos lineales, modelos polinomiales (Ljung, 1999), redes neuronales (Hippert *et al.*, 2001), etc. Independientemente del modelo escogido, sus parámetros tienen que ser calculados durante una fase de entrenamiento utilizando únicamente una parte de la serie temporal, que generalmente será la correspondiente a las muestras pasadas o conocidas.

3.2. Estrategias de predicción de series temporales

A la hora de realizar la predicción de varias muestras simultáneas de una serie temporal, es muy importante la selección de una estrategia de predicción, ya que el error dependerá en gran medida de esta. Las estrategias principales que existen a la hora de realizar la predicción son las de salida única y las de salida múltiple.

3.2.1. Estrategias de salida única

Se trata de estrategias en las que solo se predice una única muestra en cada iteración a partir de múltiples muestras ya conocidas. Sorjamaa y Lendasse (2006) muestran tres tipos de estrategias que pueden ser utilizadas para la predicción.

Recursiva

Este tipo de estrategia, también llamada iterada (Tiao y Tsay, 1994), se basa en utilizar el mismo modelo una y otra vez para realizar la predicción. Solo se predice una muestra cada vez y esa muestra predicha se usa como dato conocido para predecir la siguiente muestra, llegando un paso en el que la predicción se basa en valores que han sido predichos

previamente y no en datos reales. El método para realizar las predicciones se resume como:

$$\begin{aligned}
 \hat{x}(t+1) &= f_r[(x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(r)], \\
 \hat{x}(t+2) &= f_r[(\hat{x}(t+1), x(t), x(t-1), x(t-2), \dots, x(t-(m-2))), \theta(r)], \\
 &\vdots \\
 \hat{x}(t+s) &= f_r[(\hat{x}(t+s-1), \hat{x}(t+s-2), \dots, x(t-(m-s))), \theta(r)],
 \end{aligned} \tag{3.3}$$

donde $x(t), \dots, x(t-(m-1))$ se corresponden a los valores pasados y conocidos de la serie temporal, mientras que de $\hat{x}(t+1), \dots, \hat{x}(t+s)$ se corresponden a los valores predichos. El término $\theta(r)$ se refiere a los parámetros del modelo que conforman la función f_r . A medida que se intente aumentar el horizonte de predicción se utilizarán menos valores conocidos y más valores predichos, manteniéndose el modelo de predicción en todo momento. Esto conlleva que el error que se genera en cada predicción se vaya acumulando al realimentarse la función con estos valores erróneos.

Directa

También conocida como independiente (Kline, 2004), utiliza diferentes modelos para cada muestra a predecir, empleándose en todo momento los valores ya conocidos de la serie temporal:

$$\begin{aligned}
 \hat{x}(t+1) &= f_{d_1}[(x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(d_1)], \\
 \hat{x}(t+2) &= f_{d_2}[(x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(d_2)], \\
 &\vdots \\
 \hat{x}(t+s) &= f_{d_s}[(x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(d_s)].
 \end{aligned} \tag{3.4}$$

En esta estrategia no se genera error acumulado a lo largo de las predicciones, ya que en todo momento se están utilizando los valores conocidos. El error generado para la predicción de cada muestra es únicamente de predicción del modelo. El problema de esta estrategia es que es necesario calcular un modelo para cada muestra a predecir, por lo que se aumenta considerablemente el tiempo de cálculo, aunque con la ventaja de que se reduce el error.

DirRec

Se trata de un método híbrido de las dos estrategias anteriores (Sorjamaa y Lendasse, 2006). En este caso se utiliza para cada muestra de tiempo un modelo diferente y se introducen como variables en el modelo las muestras predichas en las iteraciones anteriores:

$$\begin{aligned}
 \hat{x}(t+1) &= f_{dr_1}[(x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(dr_1)], \\
 \hat{x}(t+2) &= f_{dr_2}[(\hat{x}(t+1), x(t), x(t-1), x(t-2), \dots, x(t-(m-1))), \theta(dr_2)], \\
 &\vdots \\
 \hat{x}(t+s) &= f_{dr_s}[(\hat{x}(t+s-1), \hat{x}(t+s-2), \dots, x(t), x(t-1), x(t-2), \dots, \\
 &\quad x(t-(m-1))), \theta(dr_s)].
 \end{aligned} \tag{3.5}$$

En cada iteración, las variables de entrada al modelo utilizado se incrementan. Estas variables se pueden incluir o no en función de si se consideran que son lo suficientemente precisas. Este método no solo aporta información acerca de la predicción, sino también acerca de la validez de las aproximaciones hechas en las iteraciones anteriores. Si no se hace una selección de las variables predichas, esta estrategia tiene los mismos problemas que las técnicas anteriores ya que se aumenta la complejidad del cálculo y algunas entradas van incrementando el error acumulado.

3.2.2. Estrategias de salida múltiple

Al contrario que las estrategias anteriores, estas tratan de predecir varias muestras en un único paso, con la idea de minimizar los errores producidos (Taieb *et al.*, 2011). Las dos principales estrategias son:

Salida Múltiple - Entrada Múltiple

La estrategia de *Multiple Input - Multiple Output* (MIMO) (Bontempi, 2008), soluciona el problema de la existencia de dependencias entre los valores estimados al tener en cuenta estos valores en la predicción y por tanto, la precisión aumenta. La estrategia MIMO calcula los valores predichos de acuerdo a

$$[\hat{x}(t+s), \dots, \hat{x}(t+1)] = f_{MIMO}[(x(t), \dots, x(t-(m-1))), \theta(MIMO)]. \quad (3.6)$$

Al igual que en los casos anteriores, el término $\theta(MIMO)$ se refiere a los parámetros del modelo que se corresponde con la función f_{MIMO} , que relaciona los valores pasados con los futuros. Esta estrategia evita los problemas de independencia que se asumen en la estrategia directa y la acumulación de errores generada por la recursiva. El problema de este sistema es que al tratar de preservar las dependencias entre los valores predichos se limita el horizonte que se puede predecir con el mismo modelo.

DIRMO

Esta estrategia trata de aglutinar los aspectos positivos de la estrategia directa y la estrategia MIMO (Taieb *et al.*, 2010). Intenta predecir un horizonte s en bloques, donde cada bloque es predicho mediante una estrategia MIMO.

La predicción de s muestras se descompone en m tareas de predicción ($m = s/r$) en las que cada salida es de un tamaño r ($r \in \{1, \dots, s\}$). Cuando el parámetro r es igual a 1, el número de iteraciones es igual a s , por lo que tenemos una estrategia de predicción directa. Por el contrario, cuando $r = s$, se trata de una estrategia MIMO. El único problema de esta estrategia es determinar el parámetro r , ya que es el que determinará la eficacia de este método. El valor de este parámetro generalmente se calcula por estimación en función de las dependencias internas de las variables de salida.

3.3. Técnicas de predicción

La predicción de series temporales ha sido estudiada en gran cantidad de campos, como pueden ser en finanzas para la predicción de los mercados, en la predicción de la transmisión de datos en redes de comunicaciones o en la predicción del consumo de energía eléctrica para la programación de la producción. Esta gran cantidad de aplicaciones ha hecho que hayan surgido multitud de técnicas para construir modelos del proceso que permitan predecir los valores futuros en base a series temporales.

Entre los muchos métodos existentes destacan aquellos basados en redes neuronales (Hippert *et al.*, 2001; Metaxiotis *et al.*, 2003), dentro de los cuales cabe resaltar los métodos que usan MLP (*MultiLayer Perceptron*) (Zhang *et al.*, 2001), sistemas expertos (Rahman y Bhatnagar, 1988), lógica borrosa (Kim y Kim, 1997) o sistemas híbridos (Amjady, 2007) entre otros.

Algunos de los métodos más utilizados y que más variantes tienen son los basados en técnicas de regresión que abarcan desde técnicas simples de regresión lineal a técnicas más complejas, como es el *Support Vector Regression* (SVR).

Modelos paramétricos

Los modelos paramétricos (Box y Jenkins, 1970) son los métodos de predicción de series temporales más sencillos, realizándose la estimación por medio de una combinación de las muestras anteriores. Son los métodos más ampliamente utilizados en toda clase de campos de aplicación. Existen varios tipos de modelos:

- **Modelo simple.** Asume que la predicción para el instante $t + 1$ es igual a la muestra en el instante t :

$$\hat{x}(t + 1) = x(t). \quad (3.7)$$

- **Modelo autorregresivo (AR).** La muestra predicha es proporcional a una serie de N muestras anteriores:

$$\hat{x}(t + 1) = c + \sum_{i=0}^{N-1} x(t - i)\varphi_i + \varepsilon(t), \quad (3.8)$$

donde φ_i son los parámetros del modelo, c es una constante y ε es ruido blanco. Los parámetros se obtienen por medio de una regresión lineal.

- **Media móvil (MA).** La muestra predicha se calcula por medio de una media móvil del ruido blanco de un proceso. Esta predicción se basa en el hecho de que una muestra está correlacionada con su antecesora por medio del error del proceso:

$$\hat{x}(t + 1) = \sum_{i=0}^{N-1} \varepsilon(t - i)\theta_i. \quad (3.9)$$

Este sistema es más complejo de implementar que el anterior, ya que es difícil determinar el ruido de un proceso.

- **Métodos ARMA.** Utilizan una combinación de los modelos anteriores para realizar la predicción. Existen varios tipos de modelos ARMA pero el más destacado y utilizado es el *Autoregressive Integrated Moving Average* (ARIMA), propuesto por Hagan y Behr (1987). El modelo generalmente se refiere como ARIMA (p, d, q) donde p, d y q son los números enteros que indican el orden de la parte autorregresiva, integral y media móvil del modelo.

General Regression Neural Network

La *General Regression Neural Network* (GRNN) (Yee y Haykin, 1999) permite la generación de estructuras continuas utilizando técnicas de interpolación multidimensional. Puede ser implementada mediante una arquitectura paralela de neuronas similares y aproxima la superficie de regresión de los datos a medida que el número de ejemplos se aproxima al infinito. Este tipo de regresión es una técnica muy utilizada en predicción de series temporales por su sencillez y eficacia. El GRNN consiste en una proyección no lineal que se calcula por

$$\hat{x}_i(t+1) = S\left(\mathbf{x}^-(t) | \mathbf{x}_i^-(t), x_i(t+1)\right) = \frac{\sum_{i=1}^N \Phi\left(\mathbf{x}^-(t) - \mathbf{x}_i^-(t)\right) x_i(t+1)}{\sum_{i=1}^N \Phi\left(\mathbf{x}^-(t) - \mathbf{x}_i^-(t)\right)}, \quad (3.10)$$

donde $\Phi(\mathbf{x}^-(t))$ es una función continua, acotada, real y simétrica en torno al origen, donde toma su máximo valor. Una de las formas más habituales es el *kernel* gaussiano de forma esférica,

$$\Phi(\mathbf{x}^-(t)) = \exp\left(-\frac{\|\mathbf{x}^-(t)\|^2}{2\sigma^2}\right), \quad (3.11)$$

que, en su forma más general, viene dada por

$$\Phi(\mathbf{x}^-(t)) = \exp\left(-\frac{\mathbf{x}^-(t)^T \mathbf{C}^{-1} \mathbf{x}^-(t)}{2}\right), \quad (3.12)$$

siendo \mathbf{C} la matriz simétrica que define un hiperelipsoide en el espacio \mathbf{x} .

Support Vector Regression

Es una técnica derivada de las *Support Vector Machines* (SVM) (Cortes y Vapnik, 1995) las cuales se usan principalmente para clasificación de datos, pero pueden ser aplicadas para problemas de regresión mediante la introducción de una función de pérdida (Müller *et al.*, 1997). Las SVR tratan de estimar una función de regresión, que relaciona un regresor de entrada $\mathbf{x}_i^-(t)$ con su salida $\hat{x}_i(t+1)$, por medio de la siguiente función lineal:

$$\hat{x}_i(t+1) = \langle \mathbf{w}, \Upsilon(\mathbf{x}_i^-(t)) \rangle + b, \quad (3.13)$$

donde b es una constante y \mathbf{w} es el vector de pesos que se obtiene mediante la minimización de un margen de error por medio de la siguiente función:

$$E = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*), \quad (3.14)$$

estando la función de minimizado sujeta a las siguientes restricciones:

$$\begin{cases} x_i(t+1) - (\langle \mathbf{w}, \Upsilon(\mathbf{x}_i^-(t)) \rangle + b) \leq \varepsilon + \xi_i^* \\ (\langle \mathbf{w}, \Upsilon(\mathbf{x}_i^-(t)) \rangle + b) - x_i(t+1) \leq \varepsilon + \xi_i \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \dots, n \end{cases} \quad (3.15)$$

La entrada $\mathbf{x}_i^-(t)$ es proyectada a un espacio de mayor dimensión por la función Υ (Cortes y Vapnik, 1995) donde se comporta de forma lineal, y ξ_i^* es la variable que marca el error superior de entrenamiento (ξ_i es el inferior) fuera de lo que se denomina tubo ε -insensible. La constante $C > 0$ es el parámetro que penaliza el error. Las restricciones impuestas implican que la mayor parte de los datos de entrada están en el tubo ε , si la muestra no está en el interior, aparece un error, que es lo que se trata de minimizar con la función. El objetivo de esta técnica es lograr una transformación de los datos de un espacio de baja dimensión, donde la función que los relaciona es no lineal, a un espacio de alta dimensión donde estos datos sí son lineales y por tanto, es más sencillo obtener la función de regresión.

Si hacemos $\varepsilon = 0$ y no se proyectan los datos en un espacio de dimensión superior, nos encontramos con el caso de una regresión por mínimos cuadrados. La función Υ puede transformar $\mathbf{x}_i^-(t)$ en un espacio de dimensión superior. Sin embargo, a la hora de resolver \mathbf{w} en un espacio de dimensión superior se trata de optimizar la siguiente función:

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T F (\alpha - \alpha^*) + \varepsilon \sum_{i=1}^N (\alpha + \alpha^*) + \sum_{i=1}^N (\alpha - \alpha^*), \quad (3.16)$$

sujeta a las restricciones:

$$\begin{cases} \sum_{i=1}^N (\alpha - \alpha^*) = 0 \\ 0 \leq \alpha_i, \quad \alpha_i^* \leq C, \quad i = 1, \dots, n \end{cases} \quad (3.17)$$

La resolución de esta ecuación determina los multiplicadores de Lagrange (α_i, α_i^*) y la función de regresión viene dada por:

$$f(\mathbf{x}_i^-(t)) = \sum_{i=1}^N (\bar{\alpha}_i - \bar{\alpha}_i^*) F(\mathbf{x}_i^-(t), \mathbf{x}^-(t)), \quad (3.18)$$

donde F es la función *kernel* que se encarga de hacer la proyección del dato $\mathbf{x}_i^-(t)$ en el espacio de alta dimensión. Se utilizan funciones *kernel* para simplificar el cálculo que pueden ser entre otras:

- **Funciones polinomiales de grado d .**

$$F(\mathbf{x}_i^-(t), \mathbf{x}^-(t)) = (\gamma \langle \mathbf{x}_i^-(t), \mathbf{x}^-(t) \rangle + 1)^d. \quad (3.19)$$

- **Funciones *Radial Basis Function Network* (RBF).**

$$F(\mathbf{x}_i^-(t), \mathbf{x}^-(t)) = \exp\left(-\frac{\|\mathbf{x}_i^-(t) - \mathbf{x}^-(t)\|^2}{\sigma^2}\right). \quad (3.20)$$

Como ya se comentó, existen otros métodos que se basan en las técnicas vistas en el capítulo 2. Así, se dispone de técnicas de predicción que utilizan un agrupamiento previo con el fin de generar modelos solo con aquellas series temporales idénticas (Cao, 2003), una gran cantidad de modificaciones del SOM que pueden ser utilizadas para predicción (Barreto, 2007) o métodos que reducen primero la dimensión para quedarse con las variables independientes (Lendasse *et al.*, 2002). Estos métodos sirven tanto para generar modelos de predicción como método previo para el tratamiento de las series temporales para luego generar modelos sobre las series procesadas.

Cada uno de estos métodos aporta una ventaja concreta pero, por norma general, son menos costosos, reducen el problema de la maldición de la dimensionalidad y permiten generar en algunos casos modelos locales, evitando el problema de los saltos bruscos. A continuación se hace un análisis más concreto de cada uno de estos métodos.

3.4. Métodos basados en *clustering*

Estos métodos se basan en la utilización de técnicas capaces de particionar los datos de entrada y generar varios modelos de predicción, en lugar de un único modelo. La partición de los datos de entrada permite resolver el problema de estacionalidad de los datos. Este problema aparece cuando los datos presentan un comportamiento dependiente del tiempo y de la época del año, por lo que al entrenar un único modelo se intentan modelar dinámicas diferentes. Al separar los datos en una serie de subconjuntos disjuntos, que agrupan los datos que presentan un comportamiento similar, y entrenar un modelo para cada subconjunto, se obtiene una mayor precisión (Milidiú *et al.*, 1999).

Estos métodos buscan crear una arquitectura de mezcla de expertos (Jacobs *et al.*, 1991) mediante la cual se divide un problema complejo en problemas más simples cuya solución se combina para obtener la solución al problema global. Los métodos basados en técnicas de *clustering* generalizan esta arquitectura en un proceso de dos etapas (Tay y Cao, 2001):

1. Una primera fase en la que se aplica una técnica de *clustering* con el fin de particionar un conjunto de datos de entrada X y separarlos en una serie de K subconjuntos compuestos por n_i datos del conjunto original. Cada grupo representa lo mejor posible las diferentes dinámicas del sistema (Cao, 2003).

$$X \rightarrow \{X_i\}_{i=1,\dots,K} \rightarrow X_i = [\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{n_i}^i]. \quad (3.21)$$

Como métodos de *clustering* utilizados están el *k-means* (Martínez Álvarez *et al.*, 2011), el cual se revisó en la sección 2.4.1, clasificadores Bayesianos e incluso redes *Fuzzy* (Abraham y Nath, 2001). La técnica más utilizada en esta primera fase es el SOM, ya que permite un agrupamiento de los datos no supervisado basándose en un aprendizaje competitivo, por lo que permite hacer una partición de los datos sin necesidad de un conocimiento previo del criterio de clasificación (Fan y Chen, 2006).

2. Una segunda fase de predicción propiamente dicha, en la que se aplican técnicas de predicción sobre los grupos obtenidos por el *clustering* anterior X_i . En esta fase se

entrenan uno o varios modelos de predicción por cada uno de los grupos obtenidos en la fase previa. Al tener los datos agrupados en función de su dinámica, se garantiza que el modelo obtenido se va a ajustar mejor para lograr la predicción.

$$X_i \rightarrow \hat{x}_i(t+1) = f_i \left[\mathbf{x}_i^-(t) \mid \Theta \right]. \quad (3.22)$$

A la hora de obtener los modelos, se usan las mismas técnicas de regresión vistas en la sección 3.3. Sin embargo, las técnicas más utilizadas son las SVR (Cao, 2003; Fan y Chen, 2006) y el k -NN (k -Nearest Neighbors) (Milidiú *et al.*, 1999).

La principal ventaja es que consiguen obtener mejores resultados que aquellos que utilizan un único modelo (Ismail *et al.*, 2011; Tay y Cao, 2001), aunque por lo general se aumenta el tiempo de procesamiento al tener que entrenar varios modelos cuyo tiempo de ejecución es elevado.

3.5. Uso del SOM para predicción

El SOM no solo puede utilizarse como método de *clustering* para agrupar series temporales con la misma dinámica, sino que también puede ser utilizado directamente para crear modelos de predicción a partir de las series temporales. Existen dos razones principales por las que se puede utilizar el SOM para predicción de series temporales (Barreto, 2007):

- **Naturaleza local de los modelos generados por el SOM.** Los modelos que genera el SOM son modelos que realizan una aproximación por medio de funciones locales, entendiéndose por locales aquellos modelos cuya salida queda determinada por operaciones matemáticas que son de aplicación solo en determinadas zonas del espacio de entrada. Modelos globales, como son los basados en redes MLP, utilizan representaciones muy distribuidas del espacio de entrada que dificultan la interpretación de los resultados. Los modelos locales permiten un mejor entendimiento de las dinámicas de los subprocesos generados por las series temporales, al menos en las regiones que se usan para calcular el modelo local. Esta propiedad del SOM, junto con su preservación de la topología, son útiles para segmentar o visualizar las series temporales (Príncipe *et al.*, 1992).
- **Arquitecturas simples de crecimiento.** El *clustering* del espacio de entrada es un paso fundamental para la eficiencia de los modelos basados en el SOM. Por tanto, es necesario obtener una buena estimación del número de vectores prototipo. Una mala estimación de la cantidad de *clusters* puede llevar a generar modelos locales que contienen pocos datos para construir un modelo correcto o tienen vectores de otras zonas con dinámicas diferentes. Una de las principales ventajas del SOM es que al ser un algoritmo no supervisado, no es necesario especificar el número de neuronas iniciales, pudiéndose utilizar técnicas de crecimiento competitivo (Fritzke, 1996).

Existen numerosas modificaciones del SOM que permiten utilizarlo de forma más eficaz en la predicción de series temporales. La mayoría de estos métodos modifican la etapa de

entrenamiento con el fin de incorporar información de la evolución temporal de la serie. En cambio, otros métodos generan modelos a partir de las neuronas entrenadas.

3.5.1. Modelo VQTAM

En la ecuación general del modelo *Vector-Quantized Temporal Associative Memory* (VQTAM) propuesto por Barreto (2007), el vector de entrada $\mathbf{x}(t)$ está compuesto por $\mathbf{x}^-(t) \in \mathbb{R}^m$ y $\mathbf{x}^+(t) \in \mathbb{R}^s$. La primera parte almacena información sobre el estado pasado de la serie temporal que va a ser aprendida, que constituye el regresor, y la segunda parte contiene información sobre el estado futuro de la serie. Teniendo esto presente, el vector prototipo de una neurona $\mathbf{w}_i(t)$ está formado por las mismas dos partes que almacenan información sobre la evolución de la variable. El algoritmo de entrenamiento del VQTAM es similar al entrenamiento del SOM, pero la neurona ganadora en un instante de tiempo t es calculada utilizando únicamente $\mathbf{x}^-(t)$:

$$c(t) = \arg \min_i \{ \|\mathbf{x}^-(t) - \mathbf{w}_i^-(t)\| \}. \quad (3.23)$$

Sin embargo, para actualizar los pesos de las neuronas, se utilizan tanto $\mathbf{x}^-(t)$ como $\mathbf{x}^+(t)$:

$$\begin{aligned} \mathbf{w}_i^-(t+1) &= \mathbf{w}_i^-(t) + \alpha(t) h_{ci}(t) [\mathbf{x}^- - \mathbf{w}_i^-(t)], \\ \mathbf{w}_i^+(t+1) &= \mathbf{w}_i^+(t) + \alpha(t) h_{ci}(t) [\mathbf{x}^+ - \mathbf{w}_i^+(t)], \end{aligned} \quad (3.24)$$

donde $0 < \alpha(t) < 1$ es la tasa de aprendizaje y $h_{ci}(t)$ es una función de vecindad gaussiana idéntica a la que se usa en el entrenamiento del SOM (ver ecuación 2.24). A medida que se va entrenando el SOM, este aprende a asociar los vectores de entrada \mathbf{w}_i^- con el correspondiente vector de salida \mathbf{w}_i^+ , preservando la topología tanto en la entrada como en la salida.

Si solo se va a predecir una única muestra (predicción MISO), los vectores $\mathbf{x}^+(t)$ y $\mathbf{w}^+(t)$ se reducen a elementos escalares $\hat{x}(t+1)$ y $w_i^+(t)$. La predicción de una muestra viene dada por

$$\hat{x}(t+1) = w_c^+(t), \quad (3.25)$$

donde la neurona ganadora se encuentra de acuerdo a la ecuación (3.23), utilizada en el entrenamiento, para cada nuevo vector de entrada \mathbf{x} .

Uno de los problemas de este método es que, al cuantizar los datos, son necesarias muchas neuronas para reducir el fallo de predicción. Para minimizar estos errores se pueden utilizar métodos de interpolación diseñados para el SOM (Yin, 2002). Los valores también pueden suavizarse utilizando redes RBF. Una red RBF, con W funciones gaussianas de base radial y una única neurona de salida, puede construirse directamente con las neuronas del SOM \mathbf{w}_i^- y \mathbf{w}_i^+ sin necesidad de entrenamiento. En este caso, la predicción para un instante de tiempo t puede determinarse como:

$$\hat{\mathbf{x}}(t+1) = \frac{\sum_{i=1}^W \mathbf{w}_i^+ G_i(\mathbf{x}^-(t))}{\sum_{i=1}^W G_i(\mathbf{x}^-(t))}, \quad (3.26)$$

siendo G_i la respuesta de esta función base al vector de entrada en el instante t ,

$$G_i(\mathbf{x}^-(t)) = \exp\left(-\frac{\|\mathbf{x}^-(t) - \mathbf{w}_i^-(t)\|^2}{2\sigma^2}\right), \quad (3.27)$$

donde \mathbf{w}_i^- funciona como centro de la i -ésima función base y $\sigma > 0$ define el radio. Este método se ha aplicado en predicción de series (Leung *et al.*, 2001; Lin y Chen, 2005).

3.5.2. SOM Recurrente

El SOM Recurrente, introducido por Koskela *et al.* (1998), incorpora integradores *leaky* a la entrada de la neurona, utilizando la recurrencia de estos integradores para capturar información que se ha perdido durante el *embedding*. Estos integradores se basan en el concepto de realimentación local y decaimiento exponencial, que actúa como un mecanismo de memoria para registrar la historia de activaciones.

Estos integradores se pueden añadir a las neuronas de la red para recordar secuencias largas a costa de una menor resolución. Utilizando estos integradores se modifica la regla de activación en el SOM, que pasa a ser

$$\mathbf{v}_i(t) = (1 - \alpha)\mathbf{v}_i(t - 1) + \alpha(\mathbf{x}(t) - \mathbf{w}_i(t)), \quad (3.28)$$

donde $\mathbf{v}_i(t)$ es el indicador de actividad de la neurona i y α determina la profundidad de la memoria a corto plazo.

Para entrenar este tipo de SOM se utilizan en la entrada una serie de muestras consecutivas, cuya longitud depende del coeficiente de *leaking*, y se determina la neurona ganadora por medio de la norma del vector \mathbf{v} : $\mathbf{v}_{BMU} = \underset{i}{\text{mín}}(\|\mathbf{v}_i(t)\|)$. Los pesos de las neuronas se actualizan mediante una variación de la regla tradicional en la que $(\mathbf{x}(t) - \mathbf{w}_i(t))$ es sustituido por $\mathbf{v}(t)$.

Existen otras variantes que utilizan recurrencia, como es el TKM (*Temporal Kohonen Map*) (Chappell y Taylor, 1993), que utiliza integradores *leaky* a la salida en vez de a la entrada. El problema de este algoritmo es que solamente mantiene la magnitud de salida en vez de capturar la dirección del error.

3.5.3. SOM Recursivo

El SOM Recursivo es muy similar al recurrente, pero mientras que este incluye recurrencia a nivel de neuronas, en el recursivo se incluye a nivel de red. El *Recursive SOM* (Voegtlin, 2002) incrementa la dimensionalidad de cada neurona al añadir un vector de pesos de salida que se compara con las salidas reales del instante anterior. El error se calcula como

$$E_i(t) = \alpha\|\mathbf{x}(t) - \mathbf{w}_i^{\mathbf{x}}(t)\|^2 + \beta\|\mathbf{v}(t - 1) - \mathbf{w}_i^{\mathbf{y}}(t)\|^2, \quad (3.29)$$

donde $\mathbf{w}_i^{\mathbf{x}}$ es el vector de pesos asociado a las series temporales y $\mathbf{w}_i^{\mathbf{y}}$ es el vector de pesos asociado a las actividades de las neuronas. El vector $\mathbf{v}(t - 1)$ recoge la actividad de las

neuronas en el estado anterior y se define del siguiente modo:

$$\mathbf{v}(t-1) = (\exp(-E_1(t-1)), \dots, \exp(-E_W(t-1))). \quad (3.30)$$

Este tipo de algoritmo es más eficiente que el SOM Recurrente al conservar información de toda la red, pero presenta un elevado coste computacional. Un modelo mucho menos exigente y más utilizado es el *Merge SOM* (Hammer *et al.*, 2004), que utiliza una representación más compacta. Este modelo almacena información acerca de la neurona ganadora en pasos anteriores por medio de la información de contexto, que se integra mediante unos vectores adicionales \mathbf{v}_i asociados a cada neurona. Como BMU se selecciona aquella que minimiza el error que mide la distancia recursiva:

$$E_i(t) = (1 - \alpha)\|\mathbf{x}(t) - \mathbf{w}_i\|^2 + \alpha\|\mathbf{v}(t) - \mathbf{v}_i\|^2, \quad (3.31)$$

donde el descriptor de contexto $\mathbf{v}(t)$ es la combinación lineal de las propiedades del ganador anterior $c(t-1)$

$$\mathbf{v}(t) = (1 - \beta)\mathbf{w}_{c(t-1)} + \beta\mathbf{v}_{c(t-1)}. \quad (3.32)$$

La regla de aprendizaje es la estándar para los pesos, pero también se actualizan los \mathbf{v}_i de la siguiente manera:

$$\Delta\mathbf{v}_i = \eta \cdot h_{ci} \cdot (\mathbf{v}(t) - \mathbf{v}_i). \quad (3.33)$$

3.5.4. Double Vector Quantization Method

El *Double Vector Quantization Method* (DVQM) es un método diseñado por Simon *et al.* (2003) que utiliza dos redes SOM:

1. Un SOM para agrupar las series temporales $\mathbf{x}^-(t)$.
2. Un SOM para agrupar las deformaciones asociadas $\Delta\mathbf{x}^-(t) = \mathbf{x}^-(t+1) - \mathbf{x}^-(t)$.

Esta doble cuantización de regresores y deformaciones solo almacena información estática de la evolución pasada de las series temporales. Sin embargo, es posible establecer una relación entre las deformaciones y su correspondiente regresor que contenga información dinámica sobre cómo las series han evolucionado entre un regresor y el siguiente. Esta relación puede modelarse por medio de una matriz de transiciones $\mathbf{P} = [p_{ij}]$, $1 \leq i \leq W_1$ y $1 \leq j \leq W_2$, donde W_1 y W_2 son los números de neuronas de los SOM entrenados y no tienen por qué ser necesariamente iguales. La matriz de probabilidades es calculada de acuerdo a la siguiente ecuación:

$$p_{ij} = \frac{|\mathbf{x}^-(t) \in C_i^{(1)} \mid \Delta\mathbf{x}^-(t) \in C_j^{(2)}|}{|\mathbf{x}^-(t) \in C_i^{(1)}|} \quad (3.34)$$

donde $C_i^{(1)}$ es la región de Voronoi correspondiente a $\mathbf{w}_i^{(1)}$ que es el i -ésimo prototipo del primer SOM, $C_j^{(2)}$ es la región de Voronoi correspondiente a $\mathbf{w}_j^{(2)}$ del segundo SOM. Esta probabilidad es una estimación de que la deformación $\Delta\mathbf{x}^-(t)$ sea representada por el prototipo $\mathbf{w}_j^{(2)}$ dado el regresor $\mathbf{x}^-(t)$, que está representado por el prototipo $\mathbf{w}_i^{(1)}$.

Algoritmo DVQM

1. Encontrar la neurona ganadora $\mathbf{w}_{c_i}^{(1)}$ en el SOM 1 para el regresor $\mathbf{x}^-(t)$.
2. Escoger una neurona $\mathbf{w}_{c_j}^{(2)}$ del SOM 2 de acuerdo a la distribución de probabilidad definida para la fila $i(t)$ de la matriz de transición \mathbf{P} .
3. Calcular una estimación del siguiente vector $\hat{\mathbf{x}}^-(t+1) = \mathbf{x}^-(t) + \mathbf{w}_{c_j}^{(2)}(t)$
4. La predicción siguiente se obtiene cogiendo la primera componente $\hat{x}_1^-(t+1)$ del vector $\hat{\mathbf{x}}^-(t+1)$ obtenido en el paso anterior.
5. Repetir los pasos 1-4 utilizando las muestras predichas como entrada para la siguiente predicción para estimar más de una muestra en el futuro (Método recursivo).

3.5.5. SOM Autorregresivos

Consisten en un entrenamiento por medio de un SOM para identificar el modelo local más apropiado para cada región a partir de los vectores prototipo. Vesanto (1997) propone un algoritmo cuyo entrenamiento es similar al del VQTAM. Una vez entrenado, se estima un modelo local para cada neurona por mínimos cuadrados Autorregresivo (AR) (m), cuyo número de parámetros es igual a la longitud de las muestras pasadas de la serie temporal. Estos modelos se entrenan con los n_i datos, del conjunto de datos de entrada N , que tienen como neurona ganadora a c_i . Existe un vector de coeficientes $\mathbf{p}_i \in \mathbb{R}^m$ asociado a cada vector prototipo \mathbf{w}_i ,

$$\mathbf{p}_i = \begin{bmatrix} p_{i,0} & p_{i,1} & \dots & p_{i,m} \end{bmatrix}^T. \quad (3.35)$$

El vector \mathbf{p}_i se calcula por medio de mínimos cuadrados,

$$\mathbf{p}_i = (\mathbf{R}_i^T \mathbf{R}_i + \lambda \mathbf{I})^{-1} \mathbf{R}_i^T \mathbf{s}_i, \quad (3.36)$$

donde $0 < \lambda \ll 1$ es una constante, \mathbf{I} es la matriz identidad y el vector \mathbf{s}_i y la matriz \mathbf{R}_i se definen como

$$\mathbf{s}_i = \begin{pmatrix} x_{1,0}^i \\ x_{2,0}^i \\ \vdots \\ x_{n_i,0}^i \end{pmatrix} \quad \text{y} \quad \mathbf{R}_i = \begin{pmatrix} 1 & x_{1,1}^i & x_{1,2}^i & \dots & x_{1,m}^i \\ 1 & x_{2,1}^i & x_{2,2}^i & \dots & x_{2,m}^i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n_i,1}^i & x_{n_i,2}^i & \dots & x_{n_i,m}^i \end{pmatrix}. \quad (3.37)$$

Con este método pueden surgir dos tipos de problemas. En primer lugar puede ocurrir que una neurona no tenga ningún dato asignado, con lo que lo mejor sería eliminar dicha neurona. En segundo lugar, que el número de vectores sea muy pequeño, con lo que el modelo no es preciso, siendo la solución unir esa neurona con sus vecinas.

Una variante de este método es el modelo *Local Linear Map* (LLM) (Walter *et al.*, 1990). Utiliza la misma idea, pero en este caso los cálculos de los modelos se realizan

Algoritmo AR-SOM

1. Entrenar un SOM con los N datos de entrada usando un modelo VQTAM. La neurona ganadora se calcula usando la parte conocida de la serie temporal (ecuación 3.23). El vector prototipo se actualiza al completo.
2. Calcular la BMU para cada dato con el que se ha entrenado el SOM. Para cada neurona se calcula un subconjunto de datos $X_i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{n_i}^i\}$ donde n_i ($0 < n_i \leq N - m$) que son los datos que tienen la neurona i como ganadora.
3. Estimar los parámetros de cada modelo local para cada subconjunto X_i de acuerdo a las ecuaciones 3.36 y 3.37.
4. Calcular la siguiente muestra de acuerdo al modelo local AR correspondiente a la neurona ganadora $\hat{x}(t+1) = \mathbf{p}_i(t) \mathbf{x}^-(t)$.

simultáneamente al entrenamiento. El número total de parámetros en el LLM es $2mW$, siendo W el número de neuronas y m el tamaño del vector de coeficientes. Una vez entrenado, la predicción es dada por el modelo local AR asociado con la neurona ganadora. El LLM actualiza los coeficientes de los vectores de los modelos locales simultáneamente con la agrupación de los vectores de datos, por lo que se reduce el tiempo invertido en la generación de los modelos.

3.5.6. KSOM

El método KSOM (Barreto *et al.*, 2004) combina la cuantización del modelo VQTAM y la capacidad de construir modelos AR locales. En lugar de trabajar con W modelos locales, este método trabaja con un único modelo AR, cuyos coeficientes son recalculados en cada instante de tiempo t a partir del subconjunto de K vectores prototipo que mejor representan la entrada. El subconjunto está compuesto por las K neuronas ganadoras:

$$\begin{aligned}
c_1(t) &= \arg \min_{\forall i} \left\{ \left\| \mathbf{x}^-(t) - \mathbf{w}_i^-(t) \right\| \right\}, \\
c_2(t) &= \arg \min_{\forall i \neq c_1} \left\{ \left\| \mathbf{x}^-(t) - \mathbf{w}_i^-(t) \right\| \right\}, \\
&\vdots \\
c_K(t) &= \arg \min_{\forall i \neq \{c_1, \dots, c_{K-1}\}} \left\{ \left\| \mathbf{x}^-(t) - \mathbf{w}_i^-(t) \right\| \right\}.
\end{aligned} \tag{3.38}$$

La predicción de la muestra para un instante de tiempo t se calcula como

$$\hat{x}(t+1) = \mathbf{p}^T(t) \mathbf{x}^-(t), \tag{3.39}$$

donde los coeficientes del vector $\mathbf{p}(t)$ son calculados mediante el método de la pseudoinversa

$$\mathbf{p}(t) = \left(\mathbf{R}^T(t) \mathbf{R}(t) + \lambda \mathbf{I} \right)^{-1} \mathbf{R}^T(t) \mathbf{s}(t), \tag{3.40}$$

donde \mathbf{I} es la matriz identidad de orden m y $\lambda > 0$ es una pequeña constante añadida a la diagonal de $\mathbf{R}^T(t)\mathbf{R}(t)$ para asegurar que la matriz es de rango máximo. El vector de predicción $\mathbf{s}(t)$ y la matriz de regresión $\mathbf{R}(t)$ se calculan como

$$\mathbf{s}(t) = \begin{pmatrix} w_{c_1,1}^+(t) \\ w_{c_2,1}^+(t) \\ \vdots \\ w_{c_K,1}^+(t) \end{pmatrix} \quad \text{y} \quad \mathbf{R}(t) = \begin{pmatrix} w_{c_1,1}^-(t) & w_{c_1,2}^-(t) & \dots & w_{c_1,m}^-(t) \\ w_{c_2,1}^-(t) & w_{c_2,2}^-(t) & \dots & w_{c_2,m}^-(t) \\ \vdots & \vdots & \ddots & \vdots \\ w_{c_K,1}^-(t) & w_{c_K,2}^-(t) & \dots & w_{c_K,m}^-(t) \end{pmatrix}. \quad (3.41)$$

El KSOM solo debe ser utilizado cuando el número de neuronas es mucho mayor que K ($W > 2K$) y únicamente está orientado a los casos en que la preservación de la topología, durante el entrenamiento no sea muy buena.

Príncipe *et al.* (1998) propone una modificación en la que el modelo local se obtiene a partir del conjunto de neuronas vecinas de la ganadora, en vez del subconjunto de K neuronas ganadoras. Para esta red, asimismo, se ha propuesto una regla de aprendizaje modificada (aprendizaje dinámico) α_ε , en la que se utiliza el error de predicción para determinar el parámetro de aprendizaje,

$$\alpha_\varepsilon = \frac{1 - \exp(-\mu(\alpha + \bar{\varepsilon}))}{1 + \exp(-\mu(\alpha + \bar{\varepsilon}))}. \quad (3.42)$$

donde μ es una constante que controla la pendiente de la curva exponencial, α es la tasa de aprendizaje convencional y $\bar{\varepsilon}$ es un valor normalizado del error que se calcula como:

$$\bar{\varepsilon} = \frac{\|\varepsilon(x(t))\|}{\|x(t+1)\|} \quad (3.43)$$

Este método es más efectivo cuando se ha conseguido una preservación de la topología durante el entrenamiento, ya que las neuronas vecinas son también las más cercanas al vector de entrada. Sin embargo, si no se ha conseguido una preservación de la topología, el KSOM genera mejores resultados.

3.6. Predicción con técnicas de *manifold learning*

Al contrario de lo que ocurre con el SOM, no existen muchas aplicaciones que utilicen otras técnicas de reducción de la dimensionalidad para predicción (Lendasse *et al.*, 2002; Chen *et al.*, 2008), debido fundamentalmente a que requieren algoritmos complejos.

La predicción con estas técnicas se basa en reducir la dimensión de los datos de entrada para generar series temporales con la evolución de los datos originales en este espacio de baja dimensión. La evolución de las variables se convierte así en una trayectoria, siendo las nuevas variables las coordenadas de las trayectorias en cada instante de tiempo. De esta forma, se consigue que el número de variables o de series generadas sea mucho menor. Además, la reducción de la dimensión garantiza que las nuevas variables (coordenadas de proyección) obtenidas en el espacio de baja dimensión sean independientes entre sí, es decir, constituyen un espacio ortogonal. Al ser independientes, cada una de las variables puede ser predicha de

forma individual, al contrario de lo que ocurre en el espacio de entrada, transformándose el problema de predecir p variables correlacionadas a predecir d variables independientes.

La predicción con estas técnicas se compone de tres pasos:

1. **Reducción de la dimensionalidad.** Se aplica un método de reducción de la dimensionalidad a los datos de entrada $\mathbb{R}^p \rightarrow \mathbb{R}^d$.

$$\begin{array}{c} \mathbb{R}^p \\ \downarrow \\ \mathbb{R}^d \end{array} \left\{ \begin{array}{l} \mathbf{x}_1^- = [x_1(t), x_1(t-1), \dots, x_1(t-(m-1))] \\ \mathbf{x}_2^- = [x_2(t), x_2(t-1), \dots, x_2(t-(m-1))] \\ \vdots \\ \mathbf{x}_p^- = [x_p(t), x_p(t-1), \dots, x_p(t-(m-1))] \end{array} \right. \quad (3.44)$$

$$\left\{ \begin{array}{l} \mathbf{y}_1^- = [y_1(t), y_1(t-1), \dots, y_1(t-(m-1))] \\ \vdots \\ \mathbf{y}_d^- = [y_d(t), y_d(t-1), \dots, y_d(t-(m-1))] \end{array} \right.$$

donde \mathbf{x}_i^- es la evolución en la trayectoria en el espacio de entrada de la coordenada i y \mathbf{y}_j^- es la evolución en el espacio de salida de la coordenada j . Una vez obtenida la nueva proyección, se genera una serie temporal con los valores de las muestras proyectadas $\{y(t)\}_{t=0}^{m-1}$, que van a generar una trayectoria, siendo el número de series independientes igual a la dimensión intrínseca d . Al igual que ocurre con cualquier regresor, existe el problema de determinar el tamaño del mismo.

2. **Predicción de la trayectoria.** Una vez contruidos los regresores hay que generar un modelo con el fin de poder predecir las coordenadas en el espacio de baja dimensión de la muestra futura $\hat{\mathbf{y}}$.

$$\left\{ \begin{array}{l} \hat{y}_1(t+1) = f_1[\mathbf{y}_1^-(t) | \Theta] \\ \vdots \\ \hat{y}_d(t+1) = f_d[\mathbf{y}_d^-(t) | \Theta] \end{array} \right. \quad (3.45)$$

A la hora de predecir estas variables se utilizan métodos clásicos como pueden ser ARIMA (Hagan y Behr, 1987), *Holt Winters* (Chatfield y Yar, 1988), modelos estructurales, etc. La predicción de las trayectorias por estos métodos tiene los mismos problemas que pueden surgir al predecir cualquier otro tipo de variable.

3. **Reconstrucción.** Una vez que por predicción se han obtenido los valores de las coordenadas en el espacio de baja dimensión $\hat{\mathbf{y}}$, hay que encontrar el punto que constituye esa proyección en el espacio de entrada $\hat{\mathbf{x}}$.

$$\begin{array}{c} \mathbb{R}^d \rightarrow \mathbb{R}^p \\ [\hat{y}_1(t+1), \dots, \hat{y}_d(t+1)] \rightarrow [\hat{x}_1(t+1), \hat{x}_2(t+1), \dots, \hat{x}_p(t+1)] \end{array} \quad (3.46)$$

En función de la técnica de reducción empleada, el método de reconstrucción diferirá. Así, por ejemplo, en el CCA la reconstrucción se basa en una estimación de distancias

en función de las distancias en la proyección, y en el LLE se basa en el cálculo de los pesos que minimizan una combinación convexa.

El problema de la reconstrucción es que nunca es perfecta ya que no se puede asegurar que el punto reconstruido se corresponda con la proyección. De entre todos los métodos de reducción de la dimensionalidad hay que usar aquellos que garantizan una reconstrucción fiable.

La mayor ventaja de estos métodos es que permiten predecir simultáneamente gran cantidad de variables minimizando el número de predicciones necesarias, bien sea porque el número de variables se reduce o porque se pueden predecir un mayor número de muestras simultáneas. Al reducir el número de iteraciones se reduce el error de predicción generado. Sin embargo, se produce una acumulación de errores al tener el error de reducción de la dimensionalidad, el error de predicción de variables propiamente dicho y el error de reconstrucción.

Existen métodos que intentan minimizar estos errores (Zhang *et al.*, 2009) que en vez de utilizar estos pasos y trabajar con una reducción de la dimensionalidad, únicamente tratan de obtener la estructura del *manifold* para modelar las series temporales y así poder predecirlas.

Estructura del sistema de medida de consumos energéticos

Para poder estudiar los datos de consumo energético, con el fin de obtener los patrones existentes y crear los modelos de evolución de dichas variables de consumo, es necesario desarrollar un sistema de medida modular y abierto que sea capaz de adquirir, almacenar, procesar y mostrar los datos de consumos energéticos de los diferentes edificios que componen el campus de la Universidad de León.

Dado que la metodología que se explica en esta tesis ha sido aplicada sobre los datos eléctricos, se describe con gran detalle la arquitectura implementada para la toma de dichos datos. Sin embargo, también se introduce la estructura para la toma de datos de otros sistemas como son los de calefacción y refrigeración, dado que dicha estructura es en gran parte común a la del sistema eléctrico.

El sistema de medida incorpora los medios necesarios para capturar la información de los distintos equipos de medida, almacenarla y procesarla por medio de una red basada en una estructura de triple capa (Domínguez *et al.*, 2005), la cual facilita la modularidad del sistema y la seguridad del mismo. Se hace necesaria la presencia de equipos que permitan el manejo y almacenamiento de todos los datos obtenidos. El sistema ha de ser capaz de presentar todos estos datos al usuario final, de tal forma que este pueda abstraerse de los detalles de bajo nivel y sea el propio sistema (de forma automática o guiado por un experto) el encargado de mostrar la información implícita en esos datos. Por este motivo, el sistema de supervisión incorporará un servidor web que haga posible el acceso remoto a los datos de consumo en los edificios de la Universidad de León.

4.1. Edificios y puntos de medida

La Universidad de León está constituida por una serie de edificios distribuidos que pueden clasificarse en tres tipos en función del uso principal que tienen:

- **Facultades y edificios de servicios.** Son edificios destinados a docencia y administración. El consumo en estos edificios es producido por sistemas de iluminación, calefacción, ventilación y aire acondicionado (*Heating, Ventilating and Air Conditio-*

ning Systems (HVAC)) (Pérez-Lombard *et al.*, 2008). Por norma general su consumo va a venir marcado fundamentalmente por el horario de clases y de trabajo del centro.

- **Edificios de investigación.** Son edificios constituidos por laboratorios y aulas. Poseen un comportamiento más de tipo industrial que destaca del resto. En estos edificios el consumo está fuertemente influenciado por equipamiento o maquinaria industrial, necesario para llevar a cabo tareas de investigación. Presentan un patrón de consumo con elevadas fluctuaciones y grandes consumos base.
- **Edificios de apoyo.** Este grupo engloba aquellos edificios que no se corresponden con ninguno de los patrones anteriores, bien sea por su uso más estacional, como puede ser el caso de una biblioteca, o bien porque tienen un consumo base mayor debido a equipamiento que trabaja las 24 horas del día, como es el caso de un centro de procesamiento de datos.

En la tabla 4.1 se enumeran los edificios que constituyen el campus de la Universidad de León y se indica dentro de qué grupo, de los anteriormente mencionados, se engloban de acuerdo a sus instalaciones.

Todos los edificios tienen instalados equipos que miden las variables de consumo, lo que permite una supervisión básica del sistema de forma local por medio de las pantallas de estos medidores. Este tipo de supervisión básica se ha de realizar de forma periódica y permite comprobar, de una forma un tanto arcaica, el consumo de cada edificio. El tiempo y los recursos requeridos en este proceso de lectura son muy elevados debido al gran número de equipos y a su distribución física dispersa por los diferentes edificios.

Lo anterior conduce a pensar en la posibilidad de definir un sistema de supervisión moderno que permita la monitorización de todas las variables de consumo devueltas, que genere históricos y sistemas de visualización, que almacene los datos de forma continua y que facilite su exportación para su tratamiento posterior. El problema de ingeniería que se plantea es diseñar un sistema de supervisión en los edificios de la Universidad de León con una estructura abierta, escalable y fiable. Esta estructura debe apoyarse en tecnologías estándar, huyendo del software propietario, debe estar preparada para la incorporación rápida de un nuevo edificio y debe ser robusta ante posibles problemas que puedan surgir.

Por lo tanto, este sistema de supervisión tiene que permitir la posibilidad de integrar herramientas tradicionales, como por ejemplo la visualización de valores instantáneos, tendencias, alarmas en base a umbrales establecidos, informes, búsqueda de valores máximos y separación de consumos en diferentes periodos, así como herramientas avanzadas de supervisión basadas en técnicas de minería de datos (Fayyad *et al.*, 1996) que aporten un sistema de visualización que permita la extracción de conocimiento de los datos de forma sencilla. Como consecuencia, la capacidad de almacenamiento de datos tiene que ser lo suficientemente grande como para no imponer un límite en la minería de datos. Un requisito vital es el acceso remoto a través de Internet. Esto hace posible la monitorización desde cualquier parte, sin necesidad de que exista un puesto central.

En el caso concreto de la medida de los consumos eléctricos, la Universidad dispone de equipos de medida tanto en los transformadores de alimentación como en cada uno de los

Facturación	Etiqueta	Edificio	Tipo de edificio
P1	E1	CRAI-TIC	Edificio de apoyo
P2	E2	Radio Universitaria	Edificio de apoyo
	E30	Enfermería	Facultad
	E31	Ciencias Económicas y Empresariales	Facultad
P3	E3	Industriales e Informática	Facultad
	E4	Aulario	Facultad
	E5	Tecnológico	Edificio de investigación
	E32	Cafetería III	Edificio de apoyo
	E17	Ciencias de la Actividad Física	Facultad
	E18	Pabellón Deportivo	Edificio de apoyo
	E19	Frontón	Edificio de apoyo
P4	E6	Filosofía y Letras	Facultad
	E7	Derecho	Facultad
	E8	Enfermería	Facultad
	E9	Ciencias del Trabajo	Facultad
P5	E10	Clínica Veterinaria	Edificio de investigación
	E11	Veterinaria	Facultad
	E12	Cafetería I	Edificio de apoyo
	E13	Ciencias Biológicas	Facultad
	E14	Servicios	Edificio de servicios
	E15	Desarrollo Ganadero	Edificio de investigación
	E16	Animalario	Edificio de investigación
P6	E20	Biblioteca San Isidoro	Edificio de apoyo
	E21	Sucursal Bancaria	Edificio de servicios
	E22	Cafetería II	Edificio de apoyo
P7	E23	Molecular	Edificio de investigación
P8	E24	Complejo Agrícolas	Edificio de investigación
P9	E25	Colegio Mayor	Edificio de apoyo
P10	E26	Complejo Rectorado	Edificio de servicios
P11	E27	Complejo la Serna	Edificio de investigación
P12	E28	Minas	Facultad
P13	E29	Centro de Idiomas	Edificio de apoyo

Tabla 4.1: Descripción de los edificios del campus de la Universidad de León.

edificios, con el fin no solo de ver el consumo de cada edificio sino también de poder comparar el consumo facturado por la compañía y el medido por el sistema de supervisión eléctrica. Estos medidores son independientes del equipo de la compañía eléctrica y son capaces de medir un amplio número de variables, a diferencia de los equipos de la compañía eléctrica, que se centran básicamente en las variables implicadas en la facturación (potencia y energías activa y reactiva).

El esquema del suministro eléctrico y los equipos eléctricos instalados en todos los edificios pueden observarse en la figura 4.1.

La compañía no factura directamente a cada edificio, sino que la Universidad tiene distintos puntos de medida. En la tabla 4.1 se recoge cómo están agrupados los edificios en función del punto de facturación eléctrica. Los edificios de la Universidad de León se pueden clasificar en dos tipos en función de la fuente de energía. Así, tenemos:

- **Edificios en media tensión.** Se alimentan directamente de la red de media tensión a 21 kV, por lo que se requiere de un transformador eléctrico que, en algunos casos es propiedad de la Universidad y, en otros casos, de la compañía eléctrica. Para estos edificios un mismo transformador alimenta a varios edificios simultáneamente, por lo que el punto de facturación de la compañía no coincide con un edificio concreto.
- **Edificios en baja tensión.** Se alimentan de la red de baja tensión a 400 V, como un edificio de oficinas cualquiera. Estos edificios corresponden a edificios situados en el centro de la ciudad. En este caso, la compañía factura directamente a cada edificio y no es necesario ningún tipo de transformador.

4.2. Arquitectura de medida y supervisión

Para poder adquirir los datos procedentes de los instrumentos de medida, almacenarlos, procesarlos y servirlos a un usuario final a través de una página web, es necesaria la creación de una estructura, tanto a nivel de software como de hardware, que permita trabajar con estos datos. Además es necesario que esta estructura cumpla ciertos requerimientos de seguridad y escalabilidad, siendo transparente al usuario final en todo momento.

Por lo general, estructuras de toma de datos de esta magnitud se encuentran en redes propias, como es el caso de la norma IEC 61850 de comunicaciones de equipos en subestaciones (Hoga y Wong, 2004), o en buses especiales de comunicación (*BACnet*). Dado que se van a incorporar a esta estructura sistemas que permitan el almacenamiento y procesamiento, es necesario la incorporación de equipos informáticos que puedan llevar a cabo estas tareas. En el caso de instalaciones simples, existen equipos que tienen integrado todo el sistema, pero para el caso de sistemas de mayor escala es necesario incorporar arquitecturas más complejas.

Para la toma de datos de consumo de la Universidad de León, se plantea la creación de una estructura multicapa que permita separar las distintas funciones necesarias en distintos niveles con funciones independientes pero que tengan relación entre sí. Existen diversos patrones, desde sistemas que solo incorporan una única capa (Nagata, 2006), a sistemas

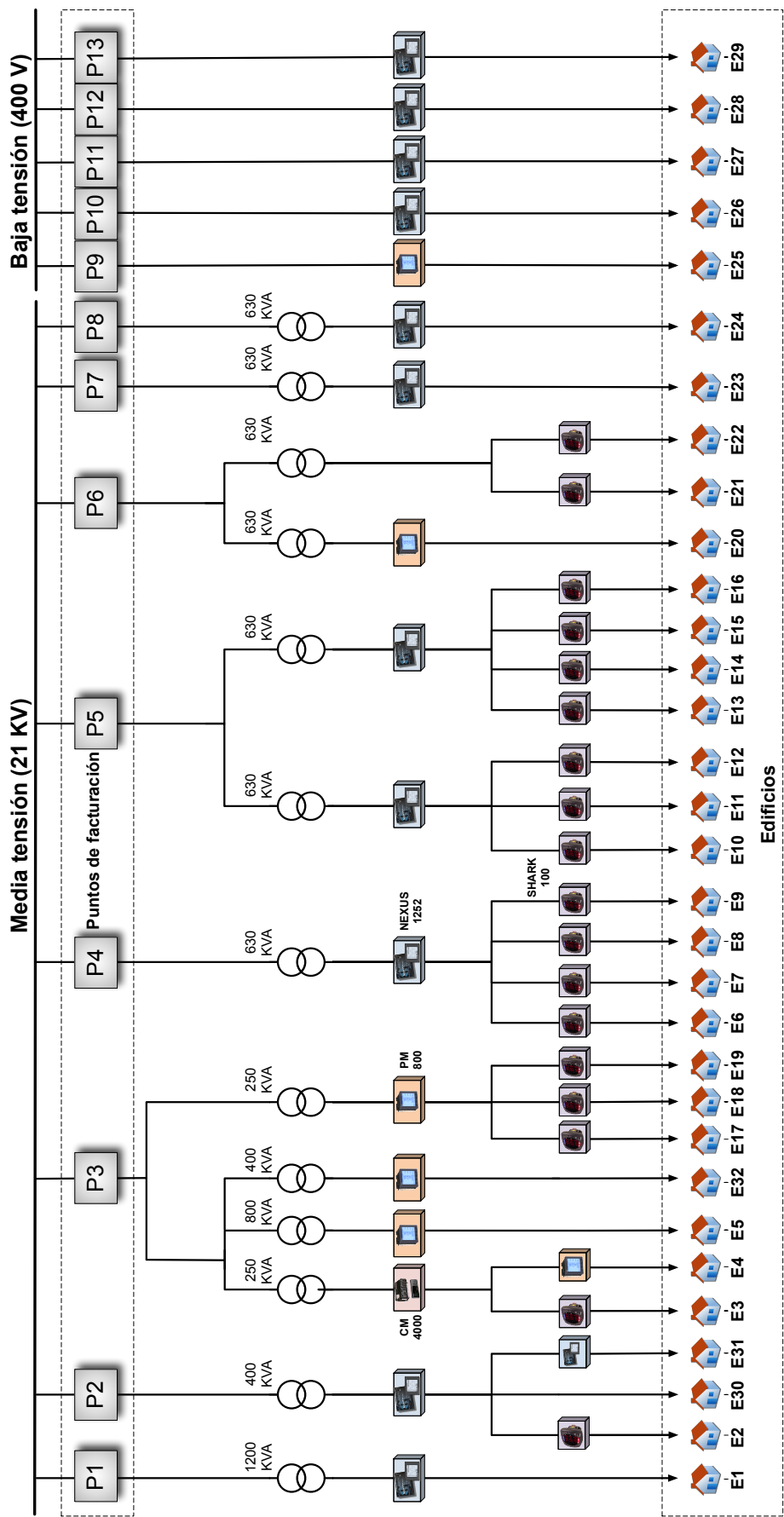


Figura 4.1: Esquema del suministro de electricidad para todos los edificios.

múltiples (Zhong *et al.*, 2005). Para la estructura del sistema de medida se adopta una arquitectura de triple capa debido al buen funcionamiento que demuestra tener en sistemas similares a nivel industrial (Eckerson, 1995; Fuertes *et al.*, 2011). Este sistema se puede asimilar a la pirámide de automatización.

Esta estructura permite establecer tres capas en las que cada una tiene su función concreta:

- **Capa física o servidor.** Es el nivel más bajo de la estructura y es donde se encuentran los equipos que realizan la captura de las variables del sistema. Todos estos equipos de medida se encuentran comunicados entre sí por medio de una red de comunicación de bajo nivel correspondiéndose con el nivel de sensores de la pirámide de automatización. Dependiendo del tipo de red utilizado, es necesaria una pasarela o tarjeta para comunicar con el siguiente nivel. En esta capa también hay que añadir conceptualmente el servidor de la compañía suministradora ya que es el que va a proporcionar los datos de las facturas.
- **Capa intermedia.** En esta capa se encuentran todos los equipos encargados del procesamiento, almacenamiento y tratamiento de los datos enviados por la capa física. Esta capa se corresponde con la capa de control de la pirámide de automatización. En función de la complejidad de los datos y de la cantidad de ellos a tratar, el número de equipos necesarios para gestionar el sistema es variable. La recomendación es distribuir las funciones entre un mayor número de equipos (Marcuse *et al.*, 1997). Estos equipos no necesariamente tienen que ser físicos, sino que es posible tenerlos todos funcionando en la misma máquina por medio de virtualizaciones (Walters, 1999). La mejor solución es tener tantos equipos como funciones:
 1. *Servidor de adquisición.* Es el servidor encargado de leer los datos procedentes de la capa física. Según el tipo de comunicación en la capa física, este servidor tiene un sistema de adquisición compatible con dicho protocolo de comunicación y envía los datos al equipo encargado del almacenamiento de los mismos.
 2. *Servidor de almacenamiento.* Este servidor posee un motor de base de datos (MySQL, MS-SQL, Oracle) en el que se almacenan tanto los datos crudos como los resultados de los distintos procesamientos realizados por el equipo encargado de ejecutar los algoritmos necesarios.
 3. *Servidor de explotación.* Este servidor lleva a cabo el tratamiento de los datos y aplica las funciones necesarias para su visualización. Los resultados son almacenados en la base de datos para que no sea necesario ejecutar los algoritmos cada vez que se consulten los datos.
 4. *Servidor web.* Es el encargado de servir los datos al usuario final. Lee los datos a mostrar directamente de la base de datos, para que en el caso de que se hagan muchos accesos al sistema, no se sature.

Dado que esta capa esta formada por sistemas informáticos estándar, la comunicación entre ellos es vía Ethernet. Puesto que los datos son enviados vía Internet, al menos el

equipo de servidor web deberá tener una IP pública. Para garantizar la seguridad del sistema y evitar posibles ataques externos, el sistema más seguro es establecer una red DMZ (*De-Militarized Zone*, Zona Desmilitarizada) (Dubrawsky *et al.*, 2006). Esta red permite aislar los equipos que han de ser accesibles desde Internet del resto de la red interna. Este sistema es parte de la red corporativa de la Universidad y actúa como *proxy* habilitando el acceso, desde el exterior, a los equipos internos del sistema.

- **Capa cliente.** Este nivel superior está formado por los equipos de los usuarios conectados al sistema. Se corresponde en la pirámide de automatización con el nivel superior que es el de supervisión. Está constituido por una red, que es Internet, a través de la cual los usuarios solicitan mediante un navegador web las interfaces de usuario para visualizar los datos. En algunos casos, en esta capa también se encuentran servidores externos al sistema, a los cuales se hacen accesos para la consulta de determinados datos suministrados por terceros, como es el caso de facturas de consumo.

En la figura 4.2 se puede ver la arquitectura del sistema de medida y supervisión de la energía eléctrica en los edificios de la Universidad de León. La figura 4.3 muestra un diagrama de la composición del sistema y de las comunicaciones entre todos los equipos que forman la estructura de triple capa.

La figura 4.2 muestra un ejemplo de comunicación a lo largo de la estructura de triple capa. A través de Internet, un usuario hace una consulta mediante una página web utilizando un equipo, que puede ser un PC o un dispositivo móvil. Una página web con gestión de usuarios permite ver los datos únicamente a los usuarios registrados. Dado que el propio servidor web actúa como proxy, habilita solo el tráfico necesario para la visualización de estos datos desde la capa cliente a los equipos internos de la red. Por lo tanto, este equipo es el que hace la consulta a la base de datos para obtener los datos necesarios para la visualización. En el caso de que el usuario sea un administrador del sistema, se puede abrir un túnel hacia la capa física a través de este servidor.

El servidor de adquisición funciona de forma independiente y periódica, dedicándose a recolectar y almacenar continuamente los datos crudos procedentes de los equipos de medida. Este servidor realiza peticiones de lectura a todos los equipos de medida y guarda los datos obtenidos en el almacén correspondiente de la base de datos. El servidor de procesamiento ejecuta los algoritmos necesarios para el procesamiento de los datos. Funciona de manera periódica, de tal forma que una serie de tareas programadas se encargan de leer los datos crudos de la base e ingresar en la misma los resultados de dichos algoritmos, donde están disponibles para una posible consulta por parte del servidor web.

A continuación, se explicará con más detalle la arquitectura del sistema de medida y supervisión desarrollada.

4.3. Capa física

La capa física está constituida por todo el equipamiento de medida tanto de las variables de consumo como de las variables ambientales, así como de los sistemas de interconexión

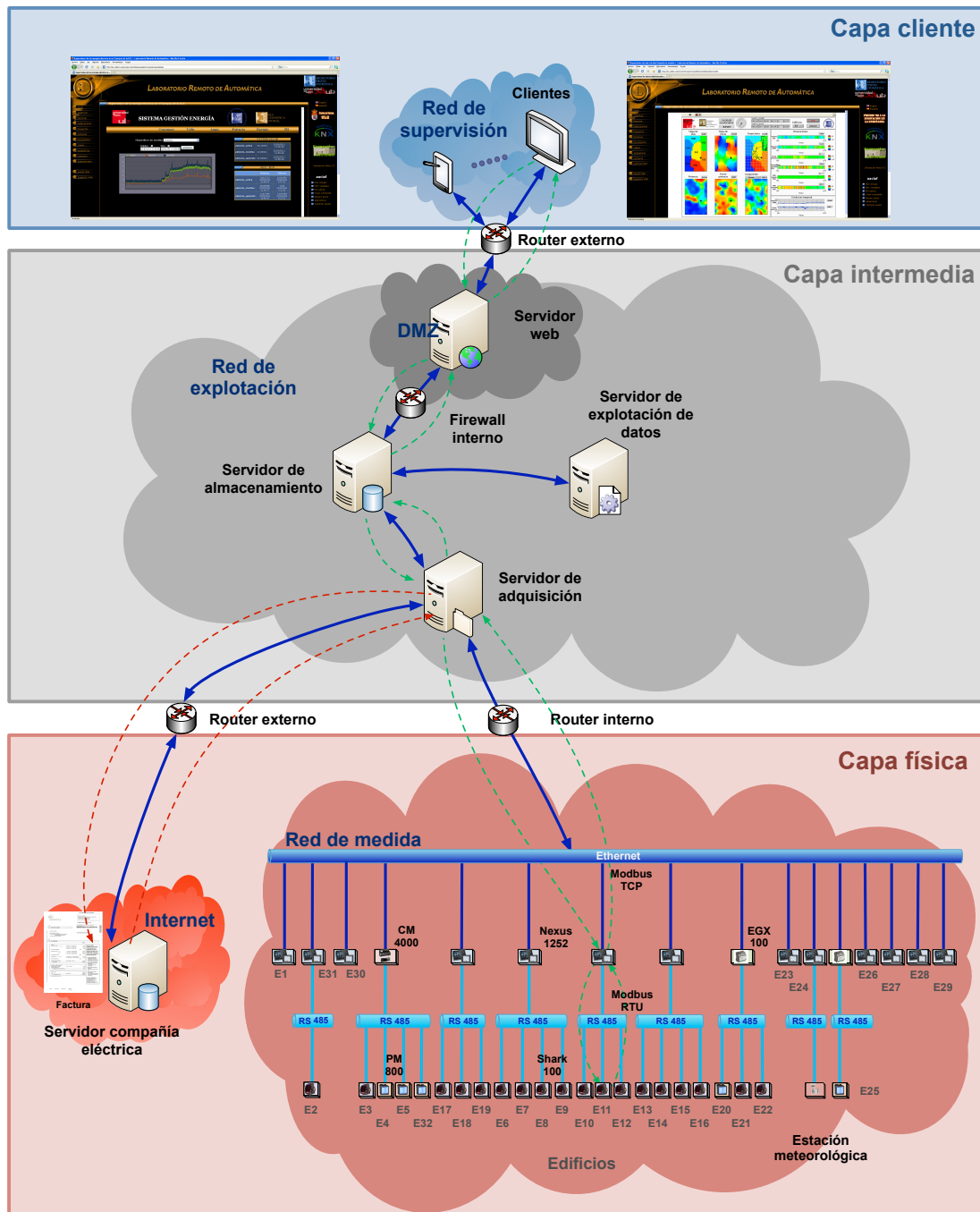


Figura 4.2: Arquitectura del sistema de medida y supervisión.

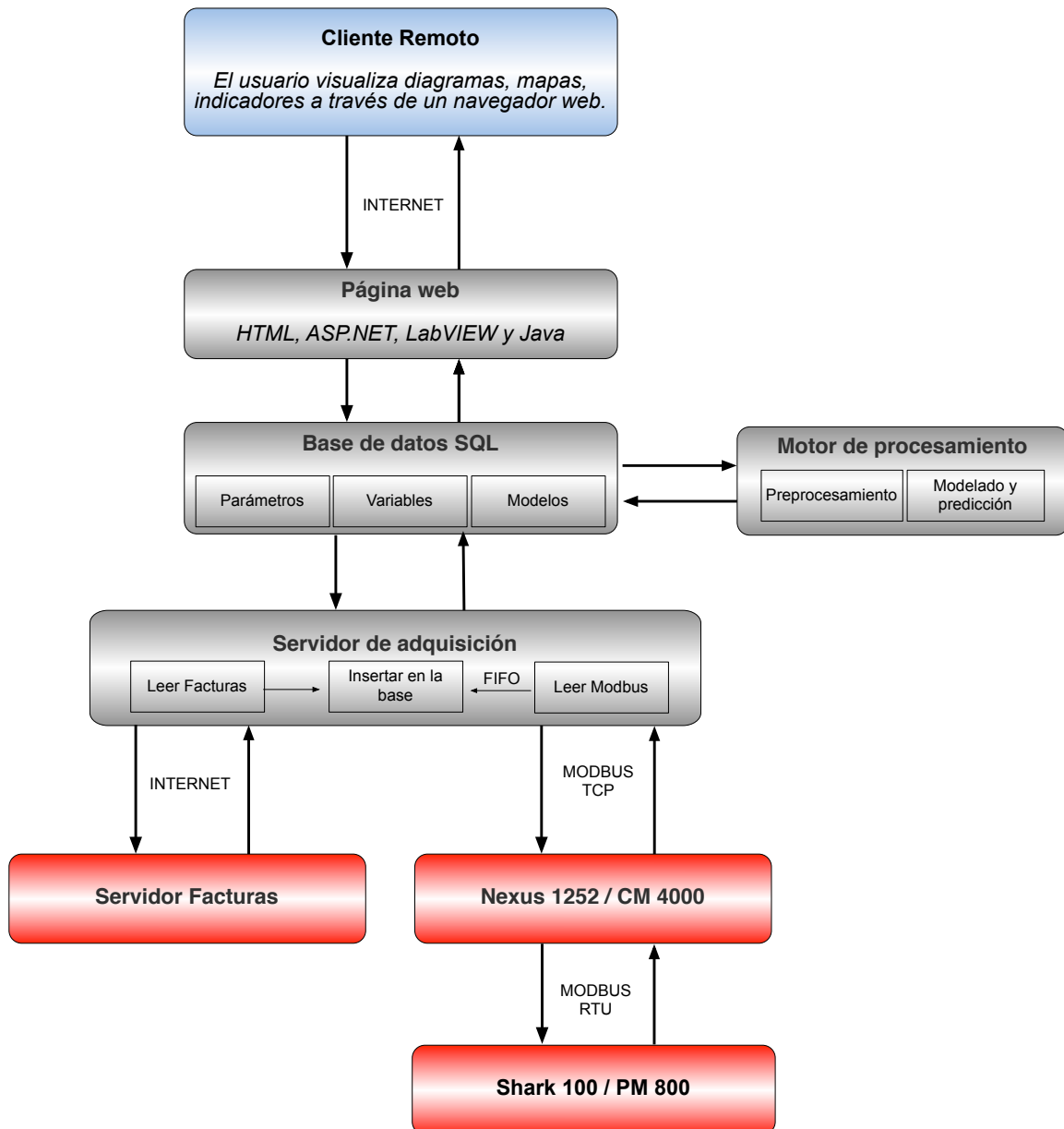


Figura 4.3: Diagrama de la arquitectura del sistema de medida y supervisión.

con la capa superior. Todo el equipamiento de esta capa se encarga de la captura de los datos por medio de una red de comunicación que interconecta todos los equipos y que le confiere una gran versatilidad y flexibilidad ante posibles averías. Este sistema de medida se puede considerar distribuido debido a la deslocalización geográfica de los edificios y a las distancias entre ellos. Por tanto, esta red establece un sistema de comunicación rápido y fiable, que enlaza todos los equipos de medida, logrando un correcto funcionamiento del sistema de medida y supervisión.

Conservando la filosofía de la triple capa, se añade un equipo a mayores en esta capa física que es el servidor de la compañía suministradora del cual se van a obtener las facturas de consumo en formato electrónico. Como se muestra en la figura 4.2, este servidor se encuentra separado del resto de la capa física, dado que no está físicamente en la misma red y es un equipo que no depende del resto de la infraestructura. Sin embargo, es un equipo que provee datos que son almacenados con el resto de datos del sistema por medio del servidor de adquisición, por lo que virtualmente está incluido en esta capa.

La capa física está formada por todos aquellos equipos que permiten adquirir las distintas variables de consumo de los edificios, así como otras variables asociadas con dicho consumo. El sistema está preparado para leer las variables eléctricas, las meteorológicas y las correspondientes a los equipos de calefacción y refrigeración, y para comunicar los equipos que toman estas variables con el servidor de adquisición bien sea por *Modbus* o por *BACnet*, que son los protocolos utilizados.

4.3.1. Equipos físicos

Equipos eléctricos

Son los equipos encargados de recoger las variables de consumo eléctrico. Están situados tanto en los puntos de facturación, donde están los transformadores eléctricos, como en la entrada de cada edificio. La comunicación interna se hace por medio de *Modbus* RTU y TCP. De esta forma es posible utilizar el medio Ethernet para la comunicación con el servidor de adquisición.

Fundamentalmente estos equipos miden tensión y corriente. El resto de posibles variables eléctricas medidas se obtienen a partir de cálculos con estas dos variables. En función de la gama del equipo y de sus funciones, el número de variables adicionales medidas por el mismo será mayor o menor. Entre las variables que pueden ofrecer los equipos eléctricos están las que se recogen en la tabla 4.2. Esta tabla muestra la magnitud del problema, ya que el número de variables hay que multiplicarlo por el número de edificios.

Existen fundamentalmente dos tipos de equipos en función de las prestaciones que ofrecen. Dentro de cada tipo de equipo existen modelos de diferentes fabricantes, lo que da peso a la necesidad de utilizar un sistema abierto para no tener que depender de software específico de los fabricantes. Los equipos utilizados son:

- **Analizadores de redes.** Están pensados para medir el consumo en la línea principal, situándose a la salida del transformador. Pueden medir gran cantidad de variables

Nº	Etiqueta	Variable	Unidad
1	Vr	Tensión entre fase r y neutro	Voltios (V)
2	Vs	Tensión entre fase s y neutro	Voltios (V)
3	Vt	Tensión entre fase t y neutro	Voltios (V)
4	Vrs	Tensión entre fases r-s	Voltios (V)
5	Vst	Tensión entre fases s-t	Voltios (V)
6	Vtr	Tensión entre fases t-r	Voltios (V)
7	Ir	Corriente en la fase r	Amperios (A)
8	Is	Corriente en la fase s	Amperios (A)
9	It	Corriente en la fase t	Amperios (A)
10	In	Corriente en el neutro	Amperios (A)
11	F	Frecuencia de red	Hercios (Hz)
12	P	Potencia activa en las tres fases	Kilovatios (kW)
13	Q	Potencia reactiva en las tres fases	Kilovoltioamperios reactivos (kVAR)
14	S	Potencia aparente en las tres fases	Kilovoltioamperios (kVA)
15	FP	Factor de potencia en las tres fases	-
16	EA	Energía activa en las tres fases	Kilovatios hora (kWh)
17	ER	Energía reactiva en las tres fases	Kilovoltioamperios reactivos hora (kVARh)
18	THD Vr	Distorsión armónica total en la tensión Vr	Porcentaje del valor fundamental (%)
19	THD Vs	Distorsión armónica total en la tensión Vs	Porcentaje del valor fundamental (%)
20	THD Vt	Distorsión armónica total en la tensión Vt	Porcentaje del valor fundamental (%)
21	THD Ir	Distorsión armónica total en la corriente Ir	Porcentaje del valor fundamental (%)
22	THD Is	Distorsión armónica total en la corriente Is	Porcentaje del valor fundamental (%)
23	THD It	Distorsión armónica total en la corriente It	Porcentaje del valor fundamental (%)
24	H1 Vr	Primer armónico en la tensión Vr	Porcentaje del valor fundamental (%)
25	H2 Vr	Segundo armónico en la tensión Vr	Porcentaje del valor fundamental (%)
26	H3 Vr	Tercer armónico en la tensión Vr	Porcentaje del valor fundamental (%)
27	H1 Ir	Primer armónico en la corriente Ir	Porcentaje del valor fundamental (%)
28	H2 Ir	Segundo armónico en la corriente Ir	Porcentaje del valor fundamental (%)
29	H3 Ir	Tercer armónico en la corriente Ir	Porcentaje del valor fundamental (%)
30	H4 Ir	Cuarto armónico en la corriente Ir	Porcentaje del valor fundamental (%)
31	H5 Ir	Quinto armónico en la corriente Ir	Porcentaje del valor fundamental (%)
32	H6 Ir	Sexto armónico en la corriente Ir	Porcentaje del valor fundamental (%)
33	H7 Ir	Séptimo armónico en la corriente Ir	Porcentaje del valor fundamental (%)
34	H1 Vs	Primer armónico en la tensión Vs	Porcentaje del valor fundamental (%)
35	H2 Vs	Segundo armónico en la tensión Vs	Porcentaje del valor fundamental (%)
36	H3 Vs	Tercer armónico en la tensión Vs	Porcentaje del valor fundamental (%)
37	H1 Is	Primer armónico en la corriente Is	Porcentaje del valor fundamental (%)
38	H2 Is	Segundo armónico en la corriente Is	Porcentaje del valor fundamental (%)
39	H3 Is	Tercer armónico en la corriente Is	Porcentaje del valor fundamental (%)
40	H4 Is	Cuarto armónico en la corriente Is	Porcentaje del valor fundamental (%)
41	H5 Is	Quinto armónico en la corriente Is	Porcentaje del valor fundamental (%)
42	H6 Is	Sexto armónico en la corriente Is	Porcentaje del valor fundamental (%)
43	H7 Is	Séptimo armónico en la corriente Is	Porcentaje del valor fundamental (%)
44	H1 Vt	Primer armónico en la tensión Vt	Porcentaje del valor fundamental (%)
45	H2 Vt	Segundo armónico en la tensión Vt	Porcentaje del valor fundamental (%)
46	H3 Vt	Tercer armónico en la tensión Vt	Porcentaje del valor fundamental (%)
47	H1 It	Primer armónico en la corriente It	Porcentaje del valor fundamental (%)
48	H2 It	Segundo armónico en la corriente It	Porcentaje del valor fundamental (%)
49	H3 It	Tercer armónico en la corriente It	Porcentaje del valor fundamental (%)
50	H4 It	Cuarto armónico en la corriente It	Porcentaje del valor fundamental (%)
51	H5 It	Quinto armónico en la corriente It	Porcentaje del valor fundamental (%)
52	H6 It	Sexto armónico en la corriente It	Porcentaje del valor fundamental (%)
53	H7 It	Séptimo armónico en la corriente It	Porcentaje del valor fundamental (%)

Tabla 4.2: Variables eléctricas capturadas por los equipos para cada edificio.

eléctricas y poseen una interfaz de comunicaciones basada en el protocolo TCP, por lo que la comunicación es más rápida y fiable. Los equipos utilizados son:

- ION 7650.
 - CM 4000.
 - Nexus 1252.
- **Medidores eléctricos.** Proporcionan un número menor de variables, muestrean a una velocidad inferior y se utilizan en los subcircuitos del sistema. Están instalados a la entrada de cada edificio, no disponen directamente de protocolo basado en TCP y solamente soportan protocolo serie, con lo que la velocidad de transmisión de información será menor y requieren de un analizador que actúe de pasarela. Los equipos utilizados son:
- Shark 100.
 - PM 800.

Equipos meteorológicos

Capturan las variables ambientales externas, que caracterizan el entorno de los edificios e influyen en el consumo eléctrico. Estas condiciones ambientales se pueden considerar comunes a todos los edificios, ya que las variaciones entre edificios van a ser mínimas debido a su proximidad geográfica. Por ello, solo se utiliza una única estación de medida, que está compuesta por los sensores necesarios para realizar las mediciones, en lugar de recurrir a una comercial, la cual es compleja de incluir en la infraestructura creada.

Para implementar la estación, se ha conectado un módulo externo de entradas analógicas a uno de los puertos RS-485 del equipo Nexus 1252. Inicialmente solo se utilizan 3 canales, uno para cada sensor conectado, parametrizados entre 4 y 20 mA. Los sensores cableados a las entradas analógicas del módulo son:

- Sensor de temperatura ambiente exterior de tipo resistivo cuyo rango de medida se encuentra entre $-35\text{ }^{\circ}\text{C}$ y $70\text{ }^{\circ}\text{C}$.
- Sensor de humedad relativa de tipo capacitivo con el rango de medida 0 % y 100 %.
- Sensor de radiación solar de tipo termoelectrónico con un rango de medida entre 0 W/m^2 y 1000 W/m^2 .

El módulo analógico adquiere continuamente los valores que proporcionan los sensores meteorológicos, mientras que el equipo Nexus 1252 registra esos datos en las correspondientes direcciones de memoria.

Equipos de calderas

Se encargan de medir las variables de consumo de los equipos de calefacción y climatización. Son recogidas por controladores específicos para el control de sistemas de

calefacción y climatización. Estos sistemas de la Universidad de León disponen de un software propietario para el control y supervisión del sistema. Sin embargo, no es abierto y utiliza sus propios equipos independientes del resto de la estructura de supervisión. Por ello, se ha utilizado software de terceros, para integrar estas medidas en la misma estructura de triple capa que se utiliza para las medidas eléctricas. Los controladores se comunican entre sí por medio de *BACnet*, utilizando este protocolo sobre TCP para la comunicación con los servidores, por lo que se puede utilizar el mismo medio físico (Ethernet) que en el caso de los equipos eléctricos para el envío de los datos a la capa intermedia.

En la Universidad, todos estos controladores son del fabricante *Siemens*, siendo de dos tipos, modulares (PXC100/200) o compactos (NCRS). La principal diferencia es que los modulares pueden ser ampliables y los compactos no.

4.3.2. Arquitectura de red

Existen varios tipos de equipos de medida que es necesario comunicar entre sí y con el servidor encargado de leer los datos registrados por estos equipos. Esta comunicación se va a realizar a través de una red Ethernet, que se denomina red de medida, la cual comparte el medio físico de la red local de la Universidad de León con el fin de minimizar la tirada de cable. Dentro de esta red de área local (*Local Area Network*, LAN), se ha definido una red de área local virtual (*Virtual Local Area Network*, VLAN) independiente y destinada únicamente al sistema de medida, la cual permite comunicar entre sí los equipos de medida energética de todos los edificios. Gracias a esta red VLAN es posible separar segmentos lógicos dentro de la red LAN de la Universidad de León y así evitar intercambiar información directamente con otros segmentos dedicados a otros fines. La red VLAN funciona en el nivel 3 o nivel de red del modelo *Open System Interconnection* (OSI) y está configurada a nivel de *routers* mediante software, creándose en estos *firewall* reglas que canalizan las comunicaciones. Para agregar nuevos equipos de medida, simplemente es necesario indicar la pertenencia de un nuevo segmento de red a la VLAN (Freeman y Passmore, 1996). Actuando de esta forma se garantiza cierta independencia y seguridad interna.

No todos los equipos van a utilizar este medio físico. Los medidores eléctricos y la periferia de los equipos de calderas se comunican por medio de cable serie RS-485 con el analizador de redes y el controlador respectivamente, los cuales hacen de pasarela con el medio Ethernet.

En cuanto a los protocolos de comunicación, existen dos, que son *Modbus* para los equipos eléctricos y meteorológicos y *BACnet* para los de calderas. Aunque los dos pueden ir sobre TCP para su transmisión vía Ethernet, son dos protocolos completamente diferentes.

Modbus

El estándar *Modbus* (Modbus Organization, 2006) es un protocolo situado en el nivel 7 o nivel de aplicación del modelo OSI. Está basado en una arquitectura maestro-esclavo en la cual el equipo responde acorde a las peticiones del maestro.

En este sistema, el servidor de adquisición de la capa intermedia actúa como maestro del sistema, siendo el único componente de la red de medida que realiza peticiones, que en

este caso son siempre de lectura. En el caso de que el equipo a acceder sea uno de los que se encuentran en alguna de las subredes, las pasarelas integradas en los analizadores de red se encargan de redirigir la petición hacia el equipo correspondiente.

Este protocolo tiene la ventaja de que es abierto, por lo que se pueden programar aplicaciones de terceros para realizar peticiones a los equipos y acceder de forma sencilla a los datos capturados por los mismos. Los campos requeridos en un paquete de datos de tipo *Modbus* son la dirección física del equipo dentro de la red, el código de la función que indica la acción a realizar por el equipo, los datos a transmitir y finalmente una comprobación de errores.

Podemos distinguir dos versiones de *Modbus*, dependiendo del medio físico sobre el cual se implementa:

- **Modbus TCP.** Esta versión de *Modbus*, donde la transmisión de los datos se realiza mediante paquetes TCP/IP, se emplea en los segmentos Ethernet. Este protocolo se utiliza en el nivel superior de la topología y en ella se encuentran el servidor de adquisición y los analizadores de redes (ION, CM y Nexus). Estos equipos permiten una comunicación directa y además llevan incorporada una pasarela que habilita la comunicación de la red de medida con el nivel inferior. Alguno de los puntos de medida no tienen instalado un analizador de redes, por lo que se incorpora una pasarela *Modbus* (EGX 100) con el fin de permitir la comunicación con el nivel inferior.
- **Modbus RTU.** Esta versión de *Modbus* se utiliza en los segmentos serie RS-485. *Modbus RTU (Remote Terminal Unit)* es una representación binaria compacta de los datos y es más eficiente que otras variantes serie, como por ejemplo *Modbus ASCII*. Este protocolo se utiliza en el nivel inferior de la topología y en ella se encuentran los medidores eléctricos (Shark, PM) y los equipos meteorológicos. Estas redes “cuelgan” de los analizadores de redes y existen varios segmentos que se corresponden con los subcircuitos que tienen los transformadores para alimentar los edificios.

BACnet

Se trata de un protocolo para el control y la automatización de edificios que fue diseñado principalmente para la comunicación y el control de sistemas tales como calefacción, climatización, iluminación, control de accesos y sistemas antiincendio (Bushby, 1997).

A diferencia de *Modbus*, este protocolo no tiene una arquitectura maestro-esclavo, sino que cada equipo publica en la red de medida las variables que lee, siendo esta transmisión accesible por todos los que conforman la red. Si el equipo es al que va destinado la trama, entonces la acepta y la procesa; en caso contrario, la rechaza. Por ello el servidor de adquisición no puede nunca pedir datos a los equipos, sino que tiene que estar escuchando continuamente la red y capturando toda la información que por ella circula.

Cuando *BACnet* se implementa sobre IP, utiliza el protocolo UDP (*User Datagram Protocol*), el cual, a diferencia del TCP que es más habitual, permite el envío de paquetes a través de Ethernet sin necesidad de establecerse previamente una conexión y sin esperar

la señal de reconocimiento de que el paquete ha sido recibido correctamente. Esto aporta como ventaja que la red se satura menos y las transmisiones son más rápidas; sin embargo, tiene el inconveniente de que no se puede saber con certeza si el dato leído es correcto.

Las tramas *BACnet* llevan información del equipo al que van destinados, del tipo de variable que se está transmitiendo y el valor de esta. Esto constituye una ventaja ya que, a la hora de capturar las variables, no es necesario configurarlas, puesto que al escuchar la red cada transmisión aporta información sobre el tipo de variable que se está capturando y es posible obtener el nombre de la misma.

Al igual que ocurría en *Modbus* existen dos medios físicos de transmisión, uno que es Ethernet y otro que es RS-485, por lo que los controladores que gestionan la red de calderas actúan como pasarela entre los dos medios. Dado que las transmisiones son lanzadas a la red, los controladores han de ser configurados como túneles para que dejen pasar las tramas desde la red RS-485 a la red Ethernet, por lo que es necesario configurar en estos cuáles son las variables que van a ser capturadas por el servidor de adquisición.

4.4. Capa intermedia

El nivel intermedio introducido entre las capas cliente y física es la capa en la que se localizan los servidores que van a llevar a cabo las tareas de adquisición, almacenamiento, procesamiento y publicación de los datos. El número de servidores dedicados a este propósito puede ser variable en función de las necesidades de cada aplicación. Para este caso, se ha optado por asignar un servidor a cada una de las funciones requeridas. A continuación se describen con más detalle los equipos y funciones de cada uno de los servidores.

4.4.1. Hardware

Aunque esta capa está formada por cuatro servidores dedicados a estas funciones, en realidad existe un único equipo físico. Dada la potencia de cálculo de los ordenadores hoy en día y que algunos de los servidores requeridos van a tener poca carga de procesamiento, la opción más económica es decantarse por un único servidor, el cual se ha virtualizado creando tantas máquinas virtuales como servidores son necesarios, aumentando con ello la seguridad y escalabilidad del sistema (Nagata, 2006).

Por tanto, el hardware en esta capa intermedia es un único equipo que va a tener dos salidas de red. Una salida es hacia la red de medida, de la cual está separada por medio de un *router*, que permite aislar la VLAN de la red de medida, y a la que se suma la red de explotación en la que se encuentran los servidores. La otra salida se direcciona hacia la capa cliente, que es directamente Internet, y por tanto va a tener una toma pública que permita a los usuarios externos acceder al equipo, que en este caso será el servidor web. Al igual que con la red de medida, existe otro *router* que separa al equipo de Internet con el fin de que actúe de cortafuegos y aumente la seguridad del equipo. El propio *router* que permite separar la capa cliente de la capa intermedia es el que sirve de enlace entre el servidor de

adquisición de datos y el servidor de la compañía suministradora que pertenece a la capa física.

El hardware utilizado es un servidor que tiene dos *sockets* con 6 núcleos cada uno (*Intel Xeon E5649* 2,53GHz) y 16 Gb de memoria RAM. Dispone de 4 discos duros de 500 Gb cada uno, montados en RAID 5 con el fin de aumentar la seguridad ante fallo, lo que le confiere una capacidad de almacenamiento total de 1,5 Tb. El servidor tiene dos tarjetas de red redundantes con dos tomas cada una en las que una toma está asignada a la red *Local Area Network* (LAN) de medida y la otra es pública. Con el fin de minimizar la posibilidad de fallo, las fuentes de alimentación también son redundantes.

La solución adoptada para la virtualización es la utilización de VMware (Oglesby y Herold, 2005). En este caso la versión gratuita aporta la suficiente funcionalidad para esta tarea. En la figura 4.4 se ve la configuración de las máquinas virtuales. Este sistema de virtualización puede asignar distintas tomas de red a cada máquina virtual, lo que permite que las máquinas estén en redes independientes a pesar de que físicamente se encuentran en el mismo equipo.

4.4.2. Software

A continuación se detallan los distintos servidores que se han implementado dentro de la máquina virtual para llevar a cabo todas las tareas sobre los datos.

Servidor de adquisición

La función de este equipo es adquirir los datos procedentes de todos los equipos de medida y enviarlos al servidor de almacenamiento, enlazando así los niveles de campo e intermedio, los cuales se encuentran físicamente en redes diferentes aunque pertenezcan a la misma red VLAN.

Este equipo está formado por una máquina virtual con un *Windows Server 2003* que tiene como recursos asignados de hardware 1 núcleo y 512 Mb de memoria RAM. Este servidor solo tiene como tarea ejecutar continuamente una serie de servicios que llevan a cabo la adquisición de datos. Se utilizan distintas aplicaciones en función de los datos que se vayan a obtener y del protocolo de comunicación que se vaya a leer:

- **Adquisición de datos eléctricos y meteorológicos vía *Modbus*.** Se ha programado un servicio en C# que realiza peticiones de lectura a todos los equipos eléctricos para obtener los datos de las variables eléctricas. Para ello, el servicio funciona como maestro de *Modbus* TCP y son los equipos de medida los que le envían datos al ser esclavos del bus y responder a la petición. De forma periódica (generalmente cada minuto, aunque este periodo constituye un parámetro que puede ser modificado), el servicio realiza una petición de datos a todos los medidores simultáneamente. Dado que se solicitan todas las variables, se envían todos los datos en bloque, al mismo tiempo, siendo el orden de recepción aleatorio, ya que depende de las demoras que se produzcan en la red. La adquisición se hace en paralelo para

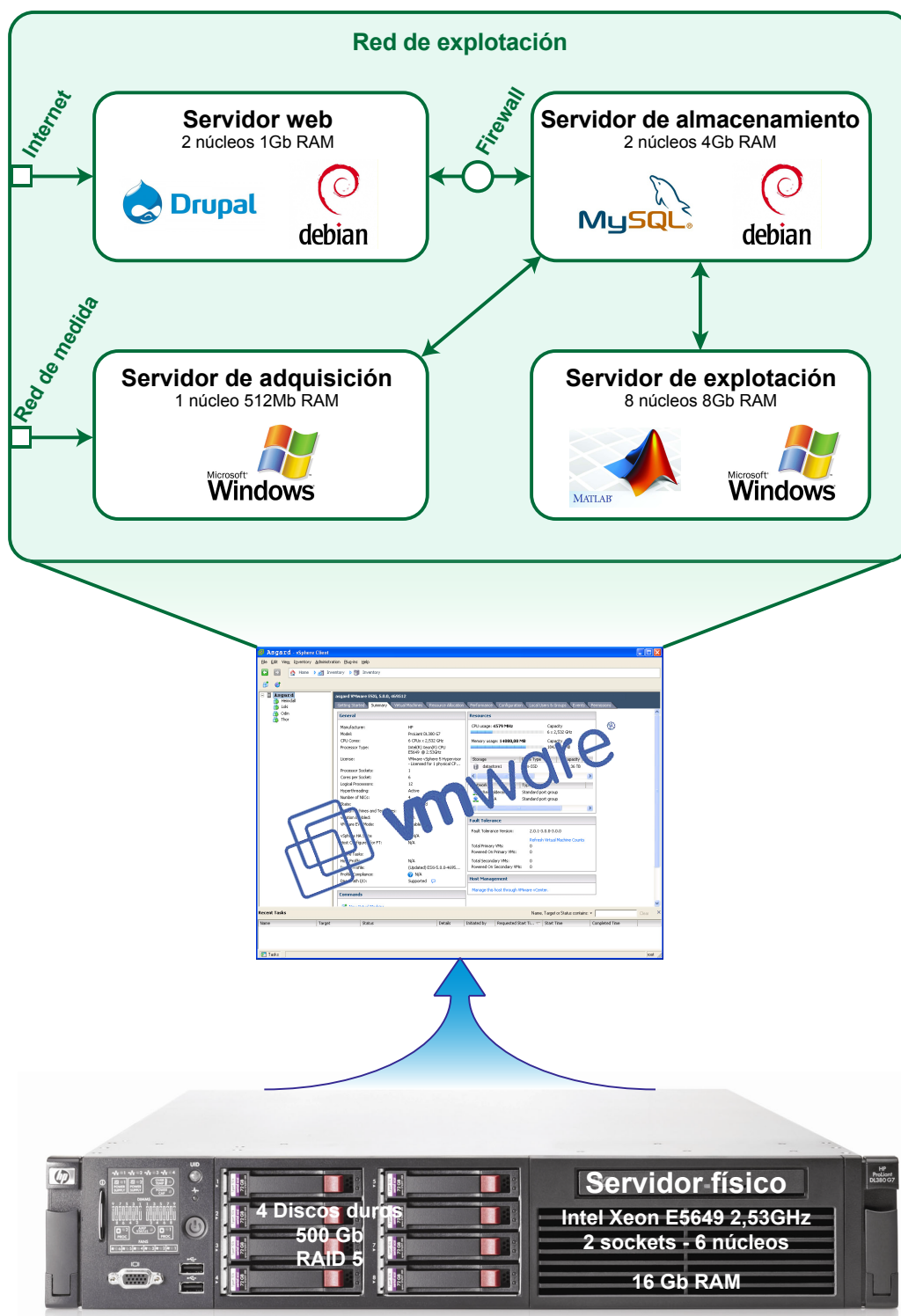


Figura 4.4: Configuración de la máquina virtual.

minimizar los retardos en la transmisión de los datos con respecto a una adquisición secuencial, donde los retardos que se producen en las peticiones a cada equipo se suman a los retardos en la respuesta. Debido al elevado número de equipos existentes en la red de medida, el número de retardos acumulado podría superar al periodo de muestreo. Una vez recopilados los datos, el servicio los inserta en la base de datos mediante un sistema de gestión de colas de tipo FIFO (*First In, First Out*), es decir, la primera muestra de datos recibida será la primera en ser insertada en la base de datos. El instante de tiempo asociado con cada muestra se corresponde con el tiempo de inserción y no con el de lectura, ya que se asume que el retardo entre la lectura y la inserción es muy inferior al periodo de muestreo.

- **Adquisición de datos de facturación de la compañía vía Internet.** También se trata de un servicio programado en C#, que accede al servidor de la compañía suministradora donde están disponibles las facturas en formato electrónico. Mediante este sistema, el servidor de adquisición es capaz de hacer accesos periódicos a este servidor para descargar la última factura. Dado que las facturas suelen ser una copia en formato digital de la factura en papel que proporciona la compañía, este programa lee esos datos en formato electrónico y realiza un análisis sintáctico con el fin de poder separar los datos importantes de la factura e insertarlos en la base de datos.
- **Adquisición de datos de calderas vía BACnet.** Gracias a la modularidad del sistema se pueden leer otro tipo de variables de consumo que son enviadas por medio de otro protocolo. En este caso se utiliza *BACnet*, y el funcionamiento de este bus es completamente diferente, ya que es necesario un sistema que no pida datos, sino que escuche continuamente el bus de comunicaciones con el fin de obtener los datos. El *BACnet*, a diferencia de *Modbus*, no responde a peticiones, sino que los datos se publican periódicamente en la red para que el sistema que necesite esos datos los recoja. Dada la complejidad para utilizar librerías que trabajen sobre *BACnet*, se ha utilizado un software propietario de terceros que es el que hace las labores de adquisición de los datos. El hecho de utilizar otro software no diseñado para esta estructura muestra la robustez de la misma, al demostrar que el sistema es completamente modular, puesto que esta aplicación simplemente lee los datos de la capa física y los inserta en la base de datos común para todo el sistema.

Dado que algunas de las aplicaciones implementadas, como se verá en la sección 4.5, requieren de un servidor web funcionando sobre *Windows*, este servidor tiene habilitado el IIS (*Internet Information Server*) con el fin de hospedar aquellas aplicaciones que lo requieran, siendo luego publicadas en la web principal por medio de procedimientos que llaman a estas aplicaciones. La razón de añadir este tipo de servicio en esta máquina virtual es el mejor aprovechamiento de los recursos asignados, ya que es la que tiene menor carga de procesador de las cuatro máquinas.

Con el fin de garantizar la seguridad del sistema, el servidor de adquisición puede tener acceso a la red Internet y a la red VLAN en la que se encuentran los equipos de medida, pudiendo así coger datos tanto de los medidores como de las facturas. Sin embargo, aunque el

equipo pueda acceder a Internet, todos los puertos de acceso desde el exterior se encuentran cerrados evitando posibles ataques, siendo el servidor web el único equipo que es accesible desde el exterior.

Servidor de almacenamiento

La función de este equipo es almacenar tanto todos los datos de consumo de los distintos equipos de medida (eléctricos, calderas y meteorológicos), como los parámetros de los modelos obtenidos. Se pueden diferenciar tres tipos de datos que pueden ser almacenados en este servidor:

- **Variables.** Constituyen tanto los datos crudos leídos por el sistema de adquisición y almacenados directamente, como los datos que ya han sido tratados y preprocesados para eliminar datos erróneos, vacantes o remuestreados, que se almacenan para evitar tener que procesarlos cada vez que se necesitan.
- **Modelos.** Son el resultado de los métodos de minería de datos que lleva a cabo el servidor de explotación. Estos serán los datos que leerán las aplicaciones de la página web y que se mostrarán al usuario final.
- **Parámetros.** Constituyen los parámetros de configuración del sistema de medida y supervisión tales como el periodo de muestreo, las direcciones físicas de los equipos dentro de la red de medida, el tipo o modelo, su permiso de activación, etc.

Este equipo es una máquina virtual sobre un sistema de tipo UNIX, más concretamente el sistema operativo es *Debian GNU/Linux*. A esta máquina se le asignan 2 núcleos y 4 Gb de RAM, puesto que las bases de datos requieren de gran cantidad de memoria RAM y se le asigna una elevada cantidad de almacenamiento con el fin de guardar los datos. Este servidor va a tener únicamente una base de datos y su sistema gestor *Database Management System* (DBMS) de tipo *MySQL*. Se utiliza esta configuración ya que tanto el sistema operativo como la base son sistemas abiertos y ampliamente utilizados, lo que evita el pago de licencias, abaratando los costes de implementación. Este equipo se encuentra en la red de explotación, por lo que la comunicación con el servidor de adquisición y el de explotación se realiza sin restricciones.

El servidor contiene distintas bases de datos de acuerdo con los tres tipos de datos (eléctricos, calderas y meteorológicos) que se pueden almacenar. Las variables, parámetros y modelos correspondientes a cada edificio se relacionan por medio de un identificador que es único, con lo que es posible obtener los modelos y los parámetros relativos a un edificio en concreto de forma sencilla.

El sistema gestor de la base de datos controla los accesos a la base por medio de la creación de distintos usuarios con funciones concretas. Hay cuatro tipos de usuarios:

- **Administración.** Son los encargados del mantenimiento de la base y tienen todos los privilegios habilitados.

- **Adquisición.** Son los usuarios mediante los cuales se accede a la aplicación de adquisición de datos y solamente puede insertar datos en la base. En ningún caso puede modificarlos.
- **Explotación.** Son usuarios que pueden leer y escribir datos pero no pueden crear tablas ni eliminar datos que no hayan creado ellos. Este tipo de usuario está pensado para poder obtener datos y crear modelos que serán almacenados. Asimismo, estos privilegios también los tiene el servidor de explotación de datos.
- **Aplicación.** Solo puede leer datos, pero en ningún caso puede modificarlos. Todos los usuarios de la capa cliente tienen este privilegio, que está pensado únicamente para poder visualizar la información almacenada en la base.

Servidor de explotación

La función de este equipo es ejecutar los programas y algoritmos empleados en el proceso de minería de datos para la obtención de las visualizaciones y modelos.

Esta máquina virtual es la que más recursos precisa, ya que algunos de los algoritmos requieren de una elevada capacidad de procesamiento para que puedan ser ejecutados en un periodo de tiempo razonable. Los recursos destinados a esta máquina son 4 núcleos y 2 *sockets*, y 8 Gb de memoria RAM. El sistema operativo es *Windows Server 2003*, aunque se podría utilizar perfectamente un sistema operativo libre como *Debian*. Los programas que se ejecutan en este servidor son programas de computación como *Matlab* y *Python*. La razón de usar *Matlab* es que la mayoría de los algoritmos de reducción de la dimensionalidad están programados y optimizados para este programa, pero con el fin de utilizar software abierto, reduciendo así el coste de implementación, estos algoritmos se pueden migrar a sistemas de software libre como son *Python* u *Octave*.

El módulo de explotación consta de 3 servicios de minería de datos diferentes, en función del objetivo perseguido:

- **Preprocesado.** Son los programas destinados al tratamiento de los datos crudos para eliminar las muestras erróneas o vacantes, *outliers* u otros valores que provocan efectos perjudiciales en etapas posteriores de la explotación de datos. Normalmente se pueden ejecutar bien por petición del usuario, para lograr un preprocesado concreto, o bien de forma periódica (por ejemplo, una vez a la semana) con el fin de preparar los datos para su uso de forma estándar y no tener que retrasar la respuesta ante una petición de procesamiento.
- **Comparación.** Son los algoritmos que usan técnicas de proyección para obtener visualizaciones de los datos con el fin, bien sea de comparar esos datos entre sí o bien de poder agrupar comportamientos similares entre los edificios. Este proceso normalmente se ejecuta de forma periódica con el fin de poder hacer un análisis del sistema, pero también se pueden ejecutar programas concretos a petición, vía web, del cliente para poder hacer análisis concretos, incluso algunos en tiempo real.

- **Modelado.** Son los programas encargados de obtener un patrón o modelo de comportamiento del edificio. Este procesado se puede realizar bien sea para mejorar los modelos obtenidos con anterioridad o para obtener modelos nuevos con los cuales realizar tareas de comparación.

Las ejecuciones de estos programas almacenados pueden realizarse por alguna de las siguientes vías:

- A petición de un usuario que quiera probar nuevas técnicas o mejorar los algoritmos ya desarrollados. En este caso se trata de usuarios locales con permisos de administrador sobre el equipo, al cual acceden generalmente por escritorio remoto para poder investigar nuevos algoritmos. Para ello, pueden recurrir a programas ya existentes.
- De forma periódica, determinada de antemano. Estas ejecuciones han sido establecidas con el fin, generalmente, de obtener visualizaciones que permitan analizar el comportamiento de los edificios a lo largo de un periodo de tiempo concreto. Se utilizan los programas ya comprobados y que no se suelen modificar.
- A petición de una aplicación web, que utiliza valores en tiempo real. Se trata de algoritmos que tratan de visualizar el comportamiento actual del edificio y que necesitan de proyecciones concretas que no pueden ser ejecutadas por la propia aplicación debido a la carencia de recursos. Por tanto, son ejecutadas en este servidor y los resultados son devueltos a la aplicación para su visualización.

Servidor web/proxy

Este equipo tiene una doble función que es, por un lado, hacer de servidor web, publicando la información para que un usuario remoto pueda supervisar el sistema, y por otro lado, actuar de servidor proxy, sirviendo de puente en aquellas situaciones en las que un usuario remoto, con privilegios, quiera acceder a alguno de los equipos internos.

Esta máquina virtual tiene como sistema operativo *Debian*, ya que la página web utilizada está basada en el gestor de contenido *Drupal*. Los recursos necesarios para esta máquina no son muy elevados, puesto que el servidor web no necesita demasiados recursos, por lo que tiene asignados 2 núcleos y 1 Gb de memoria RAM. Lo que sí es necesario es que la interfaz de red de este equipo sea una IP pública accesible desde Internet. Además, este equipo tiene que tener acceso a los otros servidores que se encuentran en la red de medida, aunque estén situados en el mismo equipo físico. Para ello, este servidor está localizado en una subred DMZ, cuyo objetivo es fortalecer la seguridad del sistema ante posibles intrusos (Dubrawsky *et al.*, 2006). De esta forma, solo se puede establecer una conexión desde la red de supervisión con el servidor web, restringiendo cualquier otra posible conexión entrante hacia la red de explotación. Para ello, se emplean dos *routers* que actúan como *firewall*, uno de ellos externo que controla el tráfico entre la red de supervisión y la subred DMZ, y otro interno que bloquea el tráfico entre la subred *De-Militarized Zone* (DMZ) y la red de explotación.

Para la función de **servidor web** se ha desarrollado un sitio web en *Drupal* (Coombs, 2009) que está programado sobre PHP. Este sitio sirve todas las aplicaciones web de supervisión y comparación implementadas, de tal forma que el cliente, cuando accede vía web puede visualizarlas. Además, tiene habilitado el acceso a través del puerto correspondiente al servidor de almacenamiento, con el fin de obtener los datos requeridos por cada una de las aplicaciones, siendo este el único puerto que está abierto de forma continua entre el servidor web y los servidores de la red de explotación. Una de las ventajas de utilizar *Drupal* es que incorpora un sistema de seguridad por el cual se puede configurar que solo los usuarios autorizados puedan acceder al contenido, dejando ciertas partes de la página ocultas al resto de usuarios. Además, también facilita la adición de contenido, al tener la posibilidad de crear nuevas páginas directamente sobre el sitio sin necesidad de conocimientos de programación.

La otra función del servidor web es la de actuar como **servidor proxy**, habilitando el acceso de determinados usuarios a equipos internos de la red. Para ello, el servidor implementa un servicio de *iptables* (Hoffman *et al.*, 2003), las cuales son reglas que permiten la apertura de puertos puntuales y, durante un periodo de tiempo, a equipos concretos internos de la red. De esta forma, un usuario registrado podría acceder al equipo de procesamiento para ejecutar determinadas aplicaciones que requieren de este servidor, o acceder a equipos de medida específicos para realizar tareas de mantenimiento, y solamente a esos. Estas reglas se crean de forma automática, por medio de un módulo PHP, a través de la página web, y son eliminadas una vez que el usuario ha salido de la página.

4.5. Capa cliente

En el nivel superior se encuadran los equipos y las herramientas utilizados por los usuarios para la visualización de los datos. Esta capa cliente es la menos definida de todas, ya que está compuesta por cualquier tipo de equipo que pueda acceder a este sistema. Dado que el servidor web está configurado siguiendo los últimos estándares, cualquier equipo que pueda ejecutar applets de Java o leer aplicaciones en HTML5 podrá ser utilizado como sistema de acceso y visualización. Por tanto, el acceso se puede hacer a través de un PC, PDA, dispositivos móviles, etc.

La red de supervisión tiene que ser accesible desde cualquier punto, permitiendo de esta forma que un usuario pueda acceder a la visualización de los datos independientemente de que se encuentre en la misma red de medida o en una red externa ajena a la Universidad. Para lograr esto, la opción más sencilla es habilitar el acceso vía Internet (Nagata, 2006).

Existen una serie de herramientas de supervisión implementadas en el servidor web programadas en distintos tipos de lenguajes de programación. Los lenguajes más utilizados son:

- **Java.** Es un lenguaje de programación ampliamente utilizado en aplicaciones web y bien conocido. Un lenguaje de programación muy utilizado para el desarrollo de herramientas de visualización es *Processing* (Reas y Fry, 2007), que es una variante de Java y permite la creación de applets que pueden ser fácilmente incrustados en el

servidor web. Las aplicaciones de Java necesitan que el usuario tenga instalada una máquina de Java para poder ejecutar los applets.

- **LabVIEW.** Es un lenguaje de programación gráfico que tiene un gran potencial en el desarrollo de interfaces gráficas (Domínguez *et al.*, 2009). Las herramientas desarrolladas en este lenguaje son muy parecidas a un SCADA tradicional ya que el propósito original de *LabVIEW* era precisamente ese. Aunque es un lenguaje muy sencillo de programar requiere de un servidor web sobre *Windows* dedicado, para poder publicar las aplicaciones. Además, el usuario necesita una máquina específica de *National Instruments* para su visualización. Las aplicaciones de *LabVIEW* se ejecutan en el servidor de adquisición, aprovechando que el equipo está infrautilizado y funciona sobre *Windows*. Mediante un *iframe*¹, a través de un puerto entre el servidor web y el de adquisición, abierto únicamente con este propósito, los paneles de *LabVIEW* son publicados en la página web.
- **HTML5/AJAX.** *Hypertext Markup Language* (HTML) versión 5 es la quinta revisión importante del lenguaje básico de la *World Wide Web*. AJAX (*Asynchronous JavaScript and XML*) es un grupo de métodos que aportan interacción en aplicaciones web (Garrett *et al.*, 2005). Este es el lenguaje en el que están programadas las páginas web. La nueva versión incorpora elementos de dibujo “canvas”, que permiten el desarrollo de herramientas de visualización interactivas similares a los *applets* de Java con la ventaja de que únicamente es necesario un navegador web que soporte este estándar, como la mayoría de los navegadores actuales (Harjono *et al.*, 2010). El principal problema de este lenguaje es que todavía es bastante reciente y es más complejo desarrollar visualizaciones vistosas.

Las aplicaciones web desarrolladas para la supervisión pueden ser clasificadas en dos grupos, atendiendo al tipo de información que se emplea en la visualización.

Herramientas tradicionales

Estas herramientas permiten la visualización de valores instantáneos, valores máximos y mínimos, tendencias de las variables, alarmas o avisos, etc. Estas herramientas gráficas pueden mostrar datos en tiempo real constituyendo una supervisión *online*. Por otra parte, también permiten recuperar datos históricos para visualizarlos y analizarlos, siendo en este caso la supervisión *offline*. Además, existe una herramienta que facilita la descarga de datos en formato texto (TXT o CSV), para su tratamiento posterior mediante una aplicación externa, utilizando por ejemplo técnicas estadísticas. En definitiva, el sistema de supervisión soporta cualquier función comúnmente utilizada por las aplicaciones de supervisión que existen en el mercado.

Se han programado una serie de herramientas tradicionales para la supervisión y el análisis de los datos:

¹Un *iframe* es un elemento HTML que permite insertar un documento HTML dentro del documento principal.



Figura 4.5: Supervisión tradicional de las facturas.

- Un entorno de supervisión desarrollado para llevar a cabo una visualización clásica de las facturas de los últimos 24 meses comparándose entre sí las de los mismos meses. En este caso se utiliza una visualización de barras idéntica a la que se suelen mostrar en las facturas enviadas por la compañía eléctrica (ver figura 4.5). Se trata de una herramienta *offline* programada en HTML5.
- Un visor de históricos de consumo, programado en ASP.NET, que muestra la evolución de una serie de variables eléctricas seleccionadas por el usuario. Este visor muestra, mediante un gráfico de líneas, los valores para un periodo de tiempo que puede ser consignado. También se trata de una herramienta *offline* ya que carece de capacidad de refresco para mostrar valores actuales.
- Una aplicación desarrollada en *LabVIEW* para la comunicación con los equipos de medida. Esta herramienta permite seleccionar el equipo para ver los valores instantáneos de las distintas variables eléctricas por lo que, a diferencia de las anteriores, se puede obtener una visualización en tiempo real. La razón de utilizar *LabVIEW* es que las herramientas desarrolladas con este lenguaje reemplazan a los programas propietarios de cada fabricante por lo que implementa un SCADA de supervisión de los equipos.

Herramientas avanzadas

Las herramientas avanzadas están basadas en las técnicas de minería de datos visual (Wong, 1999). Las visualizaciones obtenidas por estos métodos han sido incorporadas al servidor web para que puedan ser visibles a través de Internet. Al igual que las herramientas tradicionales, estas herramientas pueden ser *online*, cuando se utiliza información de datos

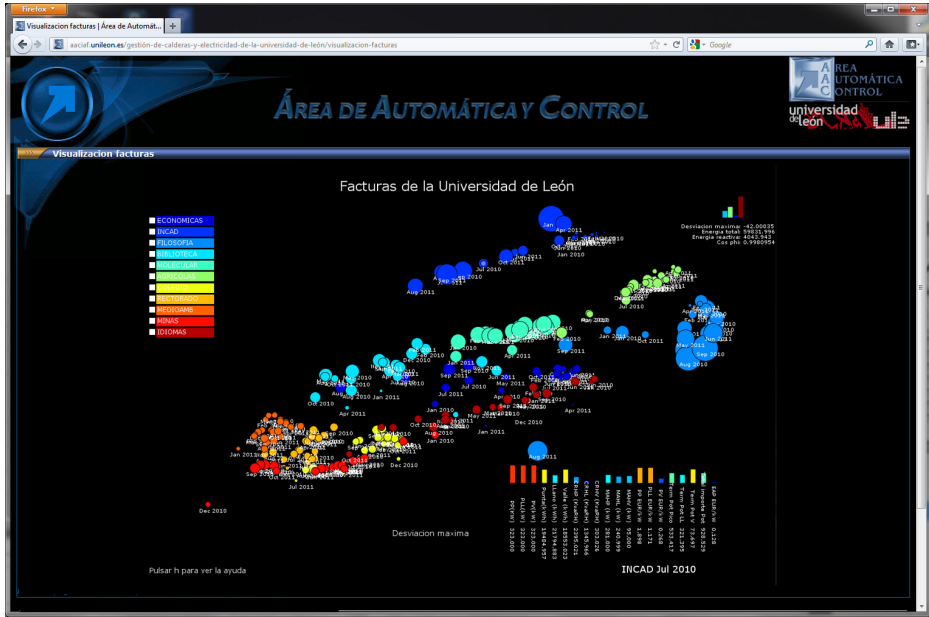


Figura 4.6: Supervisión avanzada de las facturas

actuales con las visualizaciones, pudiendo visualizarse errores, desviaciones, etc. También pueden ser herramientas *offline* cuando únicamente se quiere realizar un análisis visual de los datos con el fin de obtener información acerca de comportamientos de los edificios, visualizaciones de grupos, etc.

Las herramientas avanzadas aportan un concepto nuevo a la hora de la visualización, sobre todo porque permiten mostrar en el mismo espacio una cantidad de información mucho mayor y más ordenada que en el caso de las herramientas tradicionales. Esta visualización permite al usuario obtener conclusiones de los datos de forma más rápida y, por tanto, facilita su labor a la hora de tomar decisiones.

En el capítulo 6 se explicarán con detalle las herramientas avanzadas que han sido desarrolladas para la supervisión del sistema eléctrico de la Universidad. Estas herramientas de visualización son:

- Una herramienta para la comparación de las facturas en la que se muestran mediante puntos los distintos puntos de facturación. En la figura 4.6 se muestra el entorno desarrollado que permite la visualización simultánea de todos los puntos de facturación y una visualización más detallada de cada una de forma interactiva.
- Una herramienta para la comparación de consumos entre edificios, la cual permite visualizar las curvas diarias y su relación con las variables ambientales que condicionan el consumo. Esta herramienta aporta interactividad al usuario, permitiéndole seleccionar los puntos a visualizar y poder así comparar con más facilidad edificios o periodos de tiempo entre sí.
- Una herramienta para la estimación de los consumos energéticos de un edificio para un determinado día. Esta herramienta permite al usuario ver el posible consumo

de un edificio para unas condiciones ambientales dadas, pudiendo así planificar el consumo del campus para un determinado día del que se conocen de antemano las variables ambientales. Esta herramienta se basa en modelos que relacionan las variables ambientales con las curvas de consumo.

Todas estas herramientas son applets de Java programados en *Processing*, ya que se trata de visualizaciones interactivas y este lenguaje permite la generación de este tipo de herramientas de forma sencilla y rápida.

Metodología

Los objetivos fundamentales de esta tesis son el desarrollo de técnicas para la visualización de perfiles de consumo energético y la predicción de los mismos. Obtener una visualización que permita comparar edificios entre sí, no solo va a permitir una mejor comprensión de la distribución de los perfiles, sino que también va a permitir agrupar aquellos edificios cuya dinámica de consumo sea similar.

Esta agrupación será útil para la predicción. Crear modelos de predicción por cada edificio resulta muy costoso computacionalmente, por lo que aprovechando la agrupación de perfiles obtenida se creará un modelo por grupo. En el capítulo 3 se revisaron numerosas técnicas para predicción, por lo que se probarán varias con el fin de determinar cuál es la mejor de todas para la predicción de consumos energéticos.

5.1. Definición del espacio de entrada

El espacio de entrada está formado por un conjunto de vectores \mathbf{x} , que contienen, por un lado, información relativa a los perfiles de consumo energético, los cuales están formados por las curvas de consumo \mathbf{a} de la variable energética a estudiar para un día concreto, y, por otro, las variables ambientales \mathbf{e} para ese mismo día.

Para la creación de los perfiles se utiliza la variable que caracteriza el consumo que se va a estudiar, como puede ser la potencia activa. Para ello, se utilizan los datos correspondientes a las curvas de carga horaria y se compone un espacio de entrada en el que cada vector está formado por:

$$\mathbf{a} = \left\{ a_0, a_1, a_2, \dots, a_h \right\}, \quad (5.1)$$

donde a_h es la variable de consumo para el instante de tiempo h . Si la variable de consumo es una potencia se utiliza el valor medio a lo largo de ese instante de tiempo, si es una energía entonces se utiliza la suma. Dado que el análisis será horario, se tiene $h = 23$.

De esta forma, el espacio de entrada está formado por N vectores que contienen las curvas de consumo, una para cada día, que se obtienen de la curva de consumo anual. Cada curva de consumo lleva asociado un vector que contiene los datos de las variables ambientales medias registradas para dicho día:

$$\mathbf{e} = \left\{ \bar{e}_1, \bar{e}_2, \dots, \bar{e}_n \right\}. \quad (5.2)$$

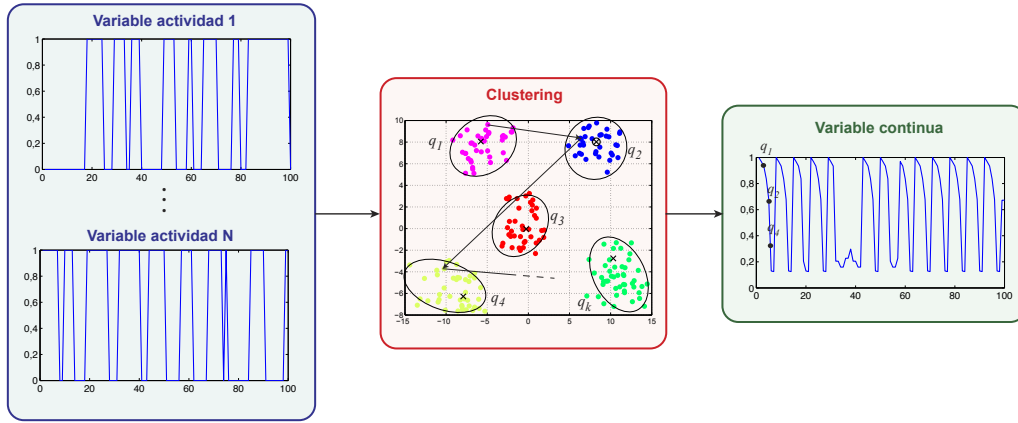


Figura 5.1: Codificación del índice de actividad.

A la hora de determinar las variables ambientales, se escogen las que tienen una influencia sobre el consumo del edificio. Por lo tanto, es necesario realizar un estudio de correlaciones entre las variables de consumo y las variables ambientales para hacer una selección de las mismas. Se requiere hacer un tratamiento de estas variables en función del tipo de dato que codifiquen. Así tenemos variables relacionadas con la meteorología, con la actividad del edificio y con el instante de tiempo:

- **Variables meteorológicas.** Estas variables, como pueden ser la temperatura o la radiación solar, son variables analógicas que generalmente tienen una correlación con la variable a analizar. Es necesario determinar la correlación entre estas variables a la hora de escoger las que se van a utilizar dentro de todas las posibles (Asbury, 1975). Este tipo de variables son continuas y no requieren procesamiento previo.
- **Variables relacionadas con la actividad del edificio.** Son variables que van a describir el tipo de día y generalmente se codifican como 0 ó 1, y que vienen dadas por descriptores del tipo de día. El uso de muchas variables binarias en métodos de proyección, como es el SOM, puede generar modelos no demasiado fiables. Además, estas variables discretas tienden a formar valores continuos en el modelo que no representan el comportamiento real de dichas variables (Lobo, 2002). En el consumo de un edificio público pueden distinguirse muchos tipos de día, como pueden ser fiestas, días laborables, puentes, etc., con lo que se pueden llegar a tener muchas variables de este tipo. La mejor solución para tratar estas variables consiste en agruparlas en la medida de lo posible por medio de técnicas de *clustering* y asignar un valor continuo q_k que las agrupe. De esta manera, se pasa de tener una serie de variables discretas a una única variable continua. Para generar este valor continuo se recurre a otro tipo de variables o indicadores que permiten valorizar la variable continua (Sideratos *et al.*, 2011). Se puede usar la potencia eléctrica, $q_k = 1/N_k \sum_{i=1}^{N_k} \bar{a}_i$, donde $i \in Q_k$, es decir, a cada grupo se le asigna el valor medio de la potencia media consumida a lo largo de cada día perteneciente a ese *cluster*. Esta valorización permite obtener una nueva variable llamada **índice de actividad**, que da una idea del uso al que está sometido el edificio (ver figura 5.1).

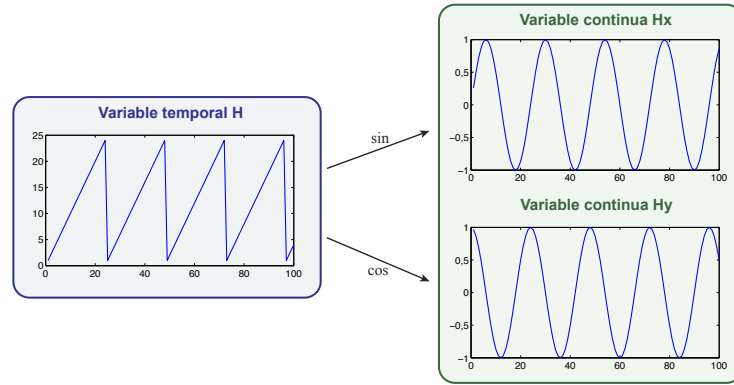


Figura 5.2: Codificación de las variables temporales.

- **Variables temporales.** Estas variables, como pueden ser el día de la semana o la hora del día, se caracterizan porque existen saltos en la evolución de la variable. Por ejemplo, en la variable hora se produce una discontinuidad al final del día (0, 23), lo que en el SOM, al usar distancias euclídeas, genera zonas alejadas en el mapa. Estas variables son interesantes porque están relacionadas con la estacionalidad de los datos, por lo que en aquellos edificios que presenten un consumo dependiente del tiempo, van a aportar información importante. Para solucionar el problema de discontinuidad, se descomponen estas variables en dos variables continuas transformadas que codifican coordenadas y que se asemejan a las manecillas de un reloj (Carpinteiro *et al.*, 2000). La transformación de estas variables es (ver figura 5.2):

$$\begin{cases} H_x = \sin\left(\frac{2\pi H}{24}\right) \\ H_y = \cos\left(\frac{2\pi H}{24}\right) \end{cases} \quad (5.3)$$

Además de esta selección de variables, un estudio de la correlación de la variable de consumo con la hora del día, con el fin de conocer la dependencia horaria de la variable es útil a la hora de generar los regresores adecuados en predicción. Esta dependencia se puede observar de forma visual, bien sea a través del análisis de las curvas de consumo o mediante el uso de una matriz de correlaciones horaria (Fay *et al.*, 2003).

Esta matriz recoge la correlación entre las horas del día para la variable de consumo y se obtiene mediante:

$$\cos(\alpha) = \frac{\sum_{i=1}^N (h_i^r - \bar{h}^r) \cdot (h_i^s - \bar{h}^s)}{\sqrt{\sum_{i=1}^N (h_i^r - \bar{h}^r)^2} \sqrt{\sum_{i=1}^N (h_i^s - \bar{h}^s)^2}}, \quad (5.4)$$

donde $\cos(\alpha)$ es el coeficiente de correlación de Pearson y h^r y h^s son los valores de la potencia para la hora r y s . Computando este valor para todas las combinaciones posibles de r y s , que toman valores $r, s = 0, \dots, 23$, se obtiene la matriz de correlación horaria. La matriz puede representarse visualmente en forma de curvas de nivel, de tal manera, que la forma

que adquieren las curvas es indicativa de la dependencia horaria. En esta representación, si las curvas son paralelas a la diagonal, entonces la dependencia horaria existente es baja, mientras que en caso contrario la dependencia es mayor (Fay *et al.*, 2003).

5.2. Proyección de los datos

Mediante técnicas de reducción de la dimensión se busca generar visualizaciones 2D que faciliten la extracción de conocimiento por parte de un usuario del comportamiento de cada edificio. Con el uso de estas técnicas de proyección se obtiene un espacio de salida en el cual existen tantos datos como en el espacio de entrada pero menor número de variables. Estas proyecciones no solo permiten determinar los distintos consumos que se pueden producir en ese edificio, sino que también posibilitan la extracción de conclusiones acerca de la relación existente entre las variables ambientales y los perfiles de consumo.

Para obtener las proyecciones de los datos se utiliza como espacio de entrada el obtenido en la sección 5.1. De esta forma, para cada punto en el espacio de salida \mathbf{y} , se tiene un vector ambiental \mathbf{e} asociado, con el cual se puede visualizar información de variables ambientales, bien sea por el color, forma o radio del punto a proyectar. A la hora de calcular la proyección, en ningún momento se considera la información aportada por los vectores ambientales, ya que lo que se busca es la proyección de los consumos \mathbf{y} , una vez que se tenga esta proyección, estudiar la relación entre las curvas de consumo y las variables ambientales. Si se calculase la proyección usando esta información, los puntos proyectados podrían distorsionarse y su proyección quedaría condicionada por la información ambiental y no dependería exclusivamente de la variable de consumo. En la figura 5.3 se muestran los pasos llevados a cabo para la creación de las proyecciones.

Proyección de los datos

1. Separar las curvas de consumo por día de la curva de consumo anual que recoge el consumo en cada instante de tiempo a lo largo de todo un año.

$$\mathbf{a} = \left\{ a_0, a_1, a_2, \dots, a_{23} \right\}$$

2. Proyectar el espacio de entrada generado por estos datos mediante una técnica de reducción de la dimensión.
3. Generar el vector ambiente, con las variables ambientales medias del día.

$$\mathbf{e} = \left\{ \bar{e}_1, \dots, \bar{e}_n \right\}$$

4. Asociar a cada vector proyectado el vector con la curva de consumo y las variables ambientales diarias. Generar la visualización donde el radio del punto es función de la variable ambiental y el color del valor medio de la potencia para ese día.

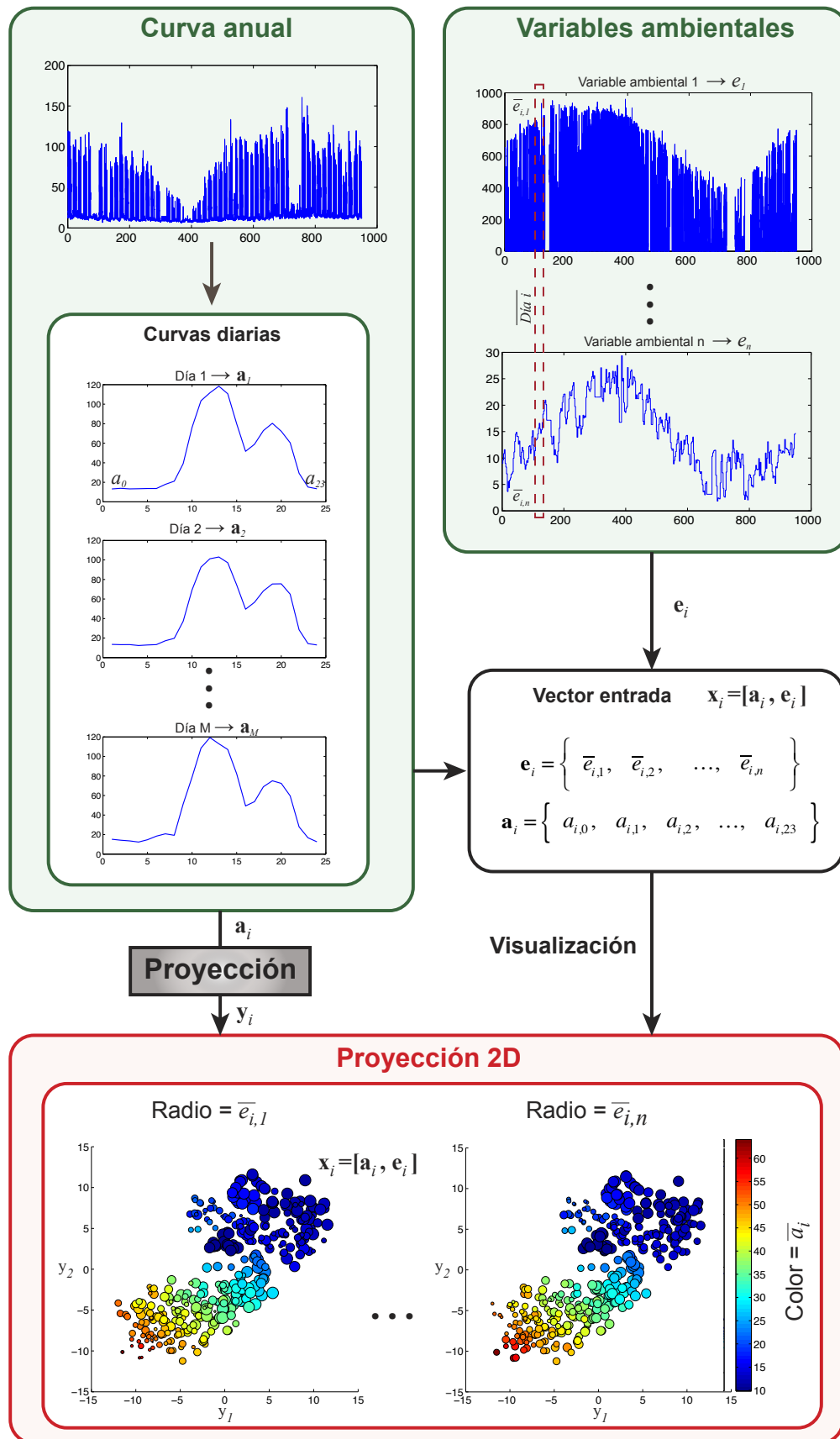


Figura 5.3: Obtención de las proyecciones de los perfiles.

En el capítulo 2 se recogen numerosas técnicas de proyección (SNE, t-SNE, CCA, CDA, Isomap, etc.), cuya idoneidad depende de la aplicación. Por lo tanto, previamente a generar las proyecciones, es necesario determinar qué método es el más efectivo. Una vez que se ha seleccionado el método y se ha generado la proyección, se ha de solventar el problema de discretización comúnmente asociado a estas técnicas, dado que no interpolan nuevos valores de los datos. Por esta razón, la proyección se mejora generando una proyección continua. Por último, se muestra una metodología para generar proyecciones que permitan no solo visualizar los perfiles de consumo de un edificio sino también poder comparar el comportamiento de consumo de los edificios entre sí.

5.2.1. Selección de la técnica

Dado que existen muchos métodos para realizar una proyección, es necesario probar varios métodos de proyección y compararlos entre sí para determinar cuál es el óptimo en el caso de los datos de perfiles energéticos. En la bibliografía no existe un acuerdo común para determinar qué método es el más efectivo para cada aplicación (Lee *et al.*, 2004; Van der Maaten *et al.*, 2007), siendo por lo general cada técnica seleccionada por el propio autor en base a su experiencia, a las características del conjunto de datos y a un análisis previo de las técnicas. Para escoger la proyección óptima se usan, fundamentalmente, los siguientes criterios:

- **Análisis visual de las proyecciones.** Dado que el principal objetivo de aplicar estas proyecciones es poder obtener conclusiones de la relación existente entre los datos, una posible forma de determinar el mejor método para la proyección de los datos es analizar directamente cuál de todos ellos es el que permite ver patrones en los datos proyectados fácilmente. En función de cómo se distribuyan los datos del espacio de entrada en el espacio de salida, se podrá obtener una conclusión (Lee y Verleysen, 2007).
- **Medida de la fiabilidad y continuidad de una visualización.** Venna y Kaski (2007); Kaski *et al.* (2003); Venna y Kaski (2006) introducen dos métricas que permiten comparar las técnicas de proyección entre sí con el fin de determinar cuál es la más óptima de forma cuantitativa. De acuerdo a su definición, se considera que una proyección es *fiable* cuando los k vecinos más cercanos de un punto en el espacio de salida son también vecinos en el espacio de entrada. Se considera N el número de muestras y $r(i, j)$ el orden de la muestra j en la lista ordenada ascendente de acuerdo a la distancia al punto i en el espacio de entrada. Sea $U_k(i)$ el conjunto de aquellas muestras que se encuentran en la vecindad de tamaño k de la muestra i en el espacio de salida, pero no en el espacio de entrada. La medida de la fiabilidad de la visualización viene dada por:

$$F(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^N \sum_{j \in U_k(i)} (r(i, j) - k). \quad (5.5)$$

Por otro lado, se considera que una proyección tiene *continuidad* cuando los k vecinos más cercanos en el espacio de entrada lo son también en el de salida. Los errores

causados por la discontinuidad pueden ser calculados de forma análoga a los errores en la fiabilidad. Consideramos $V_k(i)$ el conjunto de aquellas muestras que están en la vecindad de la muestra i en el espacio de entrada pero no en el de visualización, y sea $\hat{r}(i, j)$ el orden de la muestra j en la lista ordenada ascendente de acuerdo a la distancia desde i en el espacio de visualización. Los efectos de las discontinuidades de la proyección se miden mediante:

$$C(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^N \sum_{j \in V_k(i)} (\hat{r}(i, j) - k). \quad (5.6)$$

En el caso de que se produzcan empates a la hora de ordenar los datos, se consideran todas las ordenaciones iguales. Dado que las inicializaciones de los métodos de reducción de la dimensión son aleatorias y en algunos métodos existe un parámetro k de vecindad para entrenar el algoritmo, se deben realizar varias iteraciones con cada algoritmo y calcular el mejor y el peor caso para estos errores. Una vez calculados se considera un valor medio como indicador. Estas dos métricas se utilizan para comparar técnicas entre sí, y cuanto más cercano a 1 sea el valor de la métrica, más fiable o más continua será la proyección.

En función de los resultados de las proyecciones, tanto de forma visual como por medio de los indicadores, se seleccionará un método u otro. A pesar de la existencia de estos indicadores, es necesario también hacer un análisis visual ya que no está muy claro que los resultados de estos indicadores se correspondan con el mejor método de proyección posible. Por lo general, la técnica que se escoge es aquella que se considera que da mejor resultado, es decir, una visualización sencilla y fácilmente interpretable.

5.2.2. Mapas continuos

Cada punto proyectado en una visualización puede considerarse un estado del sistema, bien sea un proceso industrial o un análisis diario de los consumos. Uno de los principales problemas que tienen las técnicas de reducción de la dimensión es que generalmente realizan proyecciones de los puntos de manera discreta y, por tanto, no consideran posibles estados que no han tenido lugar, es decir, solo pueden almacenar información de los estados por los que ya ha pasado el proceso y no consideran la posibilidad de que ocurran nuevas situaciones en el sistema. Contemplar nuevos valores implica tener que volver a entrenar el algoritmo, por lo que los mapas generados por medio de estas proyecciones se limitan a los estados que fueron utilizados durante el entrenamiento.

Las técnicas de reducción de la dimensión no permiten realizar un mapeo inverso de forma sencilla, es decir, pasar del espacio de baja dimensión al de alta. Por ello, para reducir el coste computacional a la hora de obtener estados que no han sido proyectados en el mapa, se pueden utilizar técnicas de interpolación multidimensional que permitan obtener mapas continuos. Este problema fue considerado para el caso del SOM (Díaz, 2000; Díaz *et al.*, 2001) y para el caso de otras técnicas de reducción de la dimensión, como es el t-SNE o el Isomap (García *et al.*, 2011).

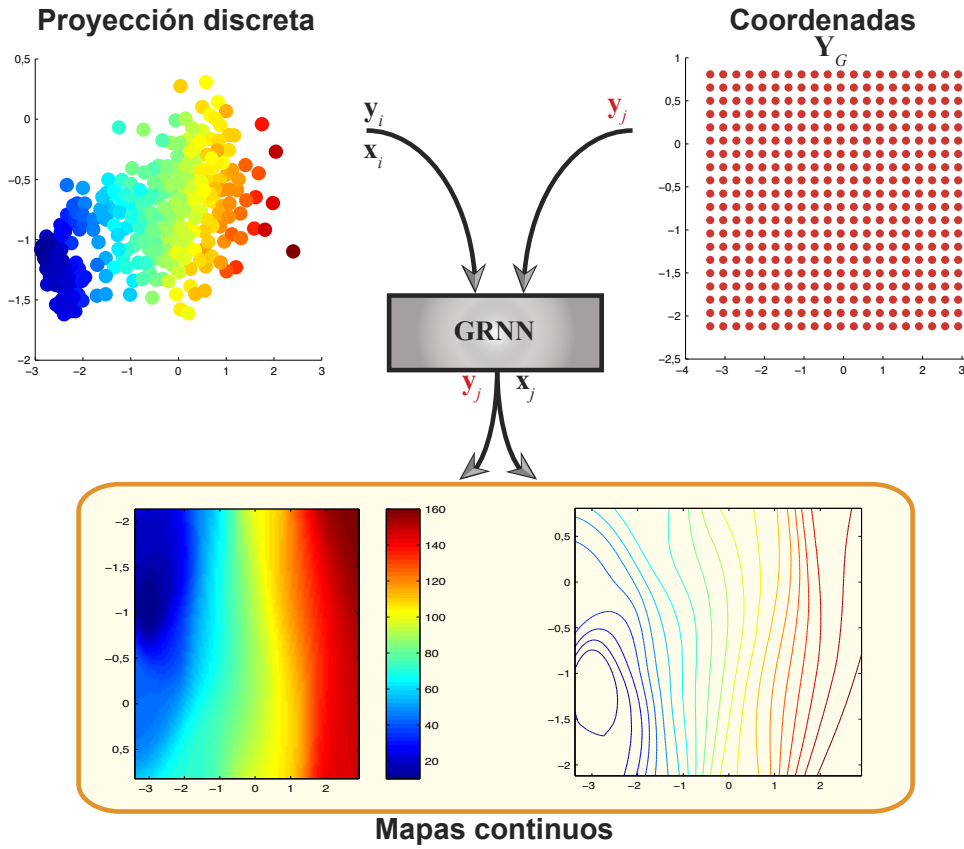


Figura 5.4: Generación de mapas continuos

Para crear estas interpolaciones se utiliza el algoritmo GRNN (Specht, 1991), que permite una proyección continua desde un espacio de elevada dimensión a uno de menor dimensión $S_{\mathbf{x}_i \rightarrow \mathbf{y}_i}$:

$$S(\mathbf{x}|\mathbf{x}_i, \mathbf{y}_i) = \frac{\sum_{i=1}^N \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2\sigma^2}\right) \mathbf{y}_i}{\sum_{i=1}^N \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2\sigma^2}\right)}. \quad (5.7)$$

Esta ecuación puede considerarse como un caso especial del Estimador de Regresión Nadaraya-Watson (NWRE) (Zaknich, 1998) que ya se vio en la sección 3.3,

$$S(\mathbf{x}|\mathbf{x}_i, \mathbf{y}_i) = \frac{\sum_{i=1}^N \Phi(\mathbf{x} - \mathbf{x}_i) \mathbf{y}_i}{\sum_{i=1}^N \Phi(\mathbf{x} - \mathbf{x}_i)}, \quad \text{donde} \quad \Phi(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right). \quad (5.8)$$

El estimador de Nadaraya-Watson puede verse como un estimador de la función $f(\mathbf{x})$

$$\mathbf{y}_i = f(\mathbf{x}_i) + \varepsilon, \quad (5.9)$$

donde la solución óptima es la esperanza de \mathbf{y} condicionada a \mathbf{x} ,

$$f(\mathbf{x}) = E[\mathbf{y}|\mathbf{x}] = \int_{R^p} \mathbf{y} p(\mathbf{y}|\mathbf{x}) d\mathbf{y} = \int_{R^p} \frac{\mathbf{y} \cdot p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})} d\mathbf{y}, \quad (5.10)$$

siendo $p(\mathbf{y}|\mathbf{x})$ la densidad de probabilidad condicional de un vector \mathbf{y} en el espacio de salida \mathbb{R}^p dado un vector \mathbf{x} en el espacio de entrada. Esta probabilidad condicional puede obtenerse mediante el conocimiento de la probabilidad conjunta de \mathbf{x} e \mathbf{y} . Esta puede ser estimada a partir de las muestras utilizando estimadores no paramétricos como el núcleo de Parzen, usado por Specht (1991). La distribución de probabilidad se calcula mediante:

$$f_h(\mathbf{x}) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \quad (5.11)$$

donde h es el parámetro de suavizado conocido como ancho de banda y $K(\cdot)$ es el núcleo que, por norma general, suele ser una distribución normal. Utilizando esta metodología es posible obtener una proyección inversa, es decir, un mapeo desde el espacio de visualización al espacio de entrada $S_{\mathbf{y}_i \rightarrow \mathbf{x}_i}$,

$$S(\mathbf{y}|\mathbf{y}_i, \mathbf{x}_i) = \frac{\sum_{i=1}^N \Phi(\mathbf{y} - \mathbf{y}_i) \mathbf{x}_i}{\sum_{i=1}^N \Phi(\mathbf{y} - \mathbf{y}_i)}. \quad (5.12)$$

Esta interpolación es la que permite generar un mapa continuo, de tal forma que mediante la creación de una rejilla que complete las coordenadas no alcanzadas por la proyección discreta se puede obtener un vector con las características del espacio de entrada para cada posición, permitiendo generar un mapa con una visualización continua. El modelo generado con GRNN permite también estimar los valores del espacio de entrada de forma interactiva, introduciendo otras posiciones fuera de la rejilla por si es requerida una mayor precisión.

Una vez obtenidos estos mapas se pueden realizar proyecciones que permitan estimar los valores de nuevas características y visualizar una distribución mediante los mapas continuos (ver figura 5.4).

Creación de mapas continuos

1. Obtener las coordenadas de la proyección de los datos \mathbf{y}_i .
2. Obtener los vectores con las curvas de consumo diarias y los valores ambientales de cada punto $\mathbf{x}_i = [\mathbf{a}_i, \mathbf{e}_i]$
3. Generar una rejilla continua \mathbf{Y}_G que contenga todos los datos proyectados.
4. Aplicar la GRNN, siendo la entrada las coordenadas de los datos proyectados y los vectores de características asociados.

$$S(\mathbf{y}|\mathbf{y}_i, \mathbf{x}_i) = \frac{\sum_{i=1}^N \Phi(\mathbf{y} - \mathbf{y}_i) \mathbf{x}_i}{\sum_{i=1}^N \Phi(\mathbf{y} - \mathbf{y}_i)}$$

5.2.3. Comparación de perfiles

Una vez que haya sido seleccionada la técnica de reducción de la dimensión, se utilizan todos los datos de los distintos edificios para proyectarlos simultáneamente y poder comparar los edificios entre sí en base a su proyección. La idea fundamental de poder proyectar todos los edificios es que, aquellos que tienen un perfil similar, aparecerán próximos en el espacio de salida. En función de la proximidad de los datos y de la distribución de estos puntos se puede estudiar la similitud de los perfiles.

A la hora de normalizar los datos (Vesanto y Alhoniemi, 2000) se puede optar por dos alternativas diferentes que generan los siguientes tipos de mapas:

- **Comparación de perfiles.** Se busca comparar los edificios en función de su consumo. Para ello se van a normalizar simultáneamente los datos al rango 0 – 1, para todos los edificios. Esta normalización se logra dividiendo todas las curvas de consumo por el punto máximo de consumo entre todos los edificios. De esta forma, al comparar edificios con valores bajos para esa variable, se proyectan distanciados de edificios que tienen un valor alto de esa variable. Para el caso particular del consumo, esta comparación permite diferenciar aquellos edificios que tienen un consumo mayor de los que consumen menos. Conocer esta diferencia entre edificios permite conocer en qué edificios es más importante actuar a la hora de ahorrar energía.
- **Comparación de dinámica de perfiles.** Se busca comparar los edificios en función de la forma de la curva de consumo. Para ello, se normalizan los datos también entre 0 – 1, dividiendo por el máximo de la variable de consumo, pero cada edificio se divide por su máximo consumo de forma independiente para conseguir que todos los edificios tras la normalización tengan un máximo de 1. Con este tipo de comparación se puede determinar qué edificios tienen una dinámica similar, lo que es útil sobre todo a la hora de poder realizar la predicción, ya que edificios con dinámica similar requieren parámetros similares.

La comparación de perfiles es principalmente visual y dado que las técnicas de reducción de la dimensionalidad generan variables ortogonales, la distribución de los puntos se va a realizar en base a variables independientes que recogen la principal información de los datos originales y que marcan las características del consumo. Dado que la distribución de los datos en el espacio de baja dimensión permite determinar los consumos similares es posible realizar un *clustering* en base a las coordenadas de los datos proyectados. Con el fin de simplificar el proceso de agrupamiento, es posible calcular la elipse que contiene a dichos datos y realizar el *clustering* en función de la forma y posición de dicha elipse, es decir, en base a la morfología de las proyecciones (Moshtaghi *et al.*, 2011; Ngan *et al.*, 2010).

5.3. Modelado de los datos

En el caso de que el número de muestras en el espacio de entrada sea muy grande, las técnicas de reducción de la dimensionalidad generan visualizaciones más complejas y

necesitan de mayor tiempo para entrenarse. El SOM es una técnica que también reduce la dimensión y permite una proyección de los datos, pero además realiza una cuantización, lo que permite reducir el número de datos. Por medio de esta técnica se genera un modelo en el que cada neurona codifica varios vectores del espacio de entrada y permite establecer una dependencia entre el consumo y el entorno ambiental (Alonso, 2012; Figueiredo *et al.*, 2005).

Para extraer conocimiento de las relaciones existentes entre variables, no es tan útil como las técnicas de proyección anteriores, ya que las visualizaciones no son tan intuitivas al ser las neuronas equidistantes unas de otras en el espacio de salida. Además, al reducir el número de datos, no se pueden llevar a cabo tareas que sí son posibles con otras proyecciones, como en el caso de estudiar la relación la curva de consumo y el día de la semana. Su ventaja fundamental es que, al reducir el número de datos, los tiempos de generación de las proyecciones son mucho más bajos y, además, permite contemplar situaciones que no han sido consideradas en el entrenamiento, al poder calcular una neurona ganadora que lo represente, mientras que en el caso de otras técnicas es necesario reentrenar la proyección para obtener la proyección de nuevos puntos.

5.3.1. Generación de modelos

Para generar los modelos con el SOM se utiliza el algoritmo de aprendizaje de Kohonen revisado en la sección 2.3.4. El espacio de entrada \mathbf{X} para el entrenamiento está compuesto, al igual que en el caso de las técnicas de proyección, por dos partes que son, por un lado, el vector que contiene la curva de consumo diaria \mathbf{a} y, por otro, las variables ambientales medias para ese día \mathbf{e} .

Sin embargo, a la hora de generar la proyección SOM, en el entrenamiento, se utiliza la información de todo el espacio de entrada, tanto la parte de consumo como las variables ambientales. Dado que se busca que estos modelos establezcan una relación entre las variables ambientales y la curva de consumo, y no tanto obtener conclusiones de forma visual acerca de las posibles relaciones existentes entre dichas variables, se obtienen mejores modelos si en el entrenamiento se utilizan las dos partes del espacio de entrada.

Además, dado que el SOM realiza una cuantización de vectores, no se puede hacer, como en el caso de las proyecciones, una asignación de vectores una vez se han proyectado los puntos, ya que el número de puntos en el espacio de salida no tiene por qué ser el mismo que en el espacio de entrada. Por tanto, el espacio de salida estará formado por una serie de coordenadas \mathbf{y}_i en el espacio \mathbb{R}^d , que se corresponden con una malla fija, y unos vectores de pesos asociados a cada neurona \mathbf{w}_i , que contienen los valores promediados de los valores del espacio de entrada \mathbf{x}_i que tienen como ganadora a esa neurona. En la figura 5.5 se detalla el proceso de generación de modelos.

El vector de pesos, al igual que el vector del espacio de entrada, está compuesto de dos componentes:

- **Parte de la curva de consumo \mathbf{w}^a .** Es la componente correspondiente a la curva de consumo \mathbf{a} de los vectores del espacio de entrada.

- **Parte ambiental \mathbf{w}^e .** Es la componente correspondiente a los valores medios diarios de las variables ambientales \mathbf{e} de los vectores del espacio de entrada.

Esta separación del vector prototipo de cada neurona permite usar una modificación en el cálculo de la neurona ganadora, de tal forma que a la hora de calcular la BMU existen tres opciones en función de la parte conocida del espacio de entrada:

- Si se conoce todo el espacio de entrada \mathbf{x} : $c = \arg \min_i \|\mathbf{x} - \mathbf{w}_i(t)\|$, que es igual que el proceso de entrenamiento (ver ecuación 2.22).
- Si solo se conoce la curva de consumo \mathbf{a} : $c = \arg \min_i \|\mathbf{a} - \mathbf{w}_i^a(t)\|$, lo cual va a permitir determinar las condiciones ambientales asociadas a una curva.
- Si solo se conocen las variables ambientales \mathbf{e} : $c = \arg \min_i \|\mathbf{e} - \mathbf{w}_i^e(t)\|$, que permitirá asociar una curva de consumo para unas determinadas condiciones ambientales.

La utilización de los modelos SOM con estos *codebooks* separados se emplean para generar visualizaciones de comparación entre edificios y también como método de predicción a largo plazo, como se verá en el apartado 5.4.

5.3.2. Comparación de modelos de perfiles

Un paso más en la comparación de modelos es requerido cuando el número de datos del espacio de entrada es grande y mientras que la comparación directa de dichos datos usando técnicas de proyección genera visualizaciones con un gran número de puntos que pueden dificultar la obtención de conclusiones por parte del usuario. De esta forma, por ejemplo, cuando se quieren comparar datos de un único año, la técnica de comparación vista en 5.2.3 es útil, pero para comparaciones de más de un año, el número de datos puede ser demasiado elevado y es necesaria una cuantización previa, que se puede lograr con un modelo generado con el SOM. Por este motivo, para comparar los modelos obtenidos con el SOM se propone el uso de los mapas de diferencias introducidos por Fuertes *et al.* (2005) y Domínguez *et al.* (2012b). Estos mapas permiten comparar dos modelos, al comparar entre sí aquellas neuronas similares, mostrando la diferencia existente para condiciones similares.

Sean dos modelos A y B , correspondientes a los perfiles de dos edificios o al perfil de un edificio en dos periodos de tiempo diferentes, de forma que $\{\mathbf{w}_i^{(A)}\}_{i=1\dots W_1}$ y $\{\mathbf{g}_i^{(A)}\}_{i=1\dots W_1}$ son los vectores de codificación y las posiciones de la retícula neuronal en la estructura SOM del modelo A , y $\{\mathbf{w}_j^{(B)}\}_{j=1\dots W_2}$ y $\{\mathbf{g}_j^{(B)}\}_{j=1\dots W_2}$ lo son del modelo B . Se define entonces la **función vectorial diferencia entre modelos**, $D_{AB} : \mathbb{R}^2 \rightarrow \mathbb{R}^n$, como la que asigna a cada neurona $\mathbf{y}_j^{(B)}$ del modelo B la diferencia entre su vector prototipo $\mathbf{w}_j^{(B)}$ en el espacio de entrada y el vector prototipo BMU $\mathbf{w}_{c_j}^{(A)}$ de esta neurona en el modelo A . $\mathbf{w}_{c_j}^{(A)}$ representa el estado del modelo A con mayor afinidad al estado del modelo B .

El criterio de selección de la BMU es el de la más próxima según una cierta función de distancia,

$$c_j = \arg \min_i \|\mathbf{w}_j^{(B)} - \mathbf{w}_i^{(A)}\|. \quad (5.13)$$

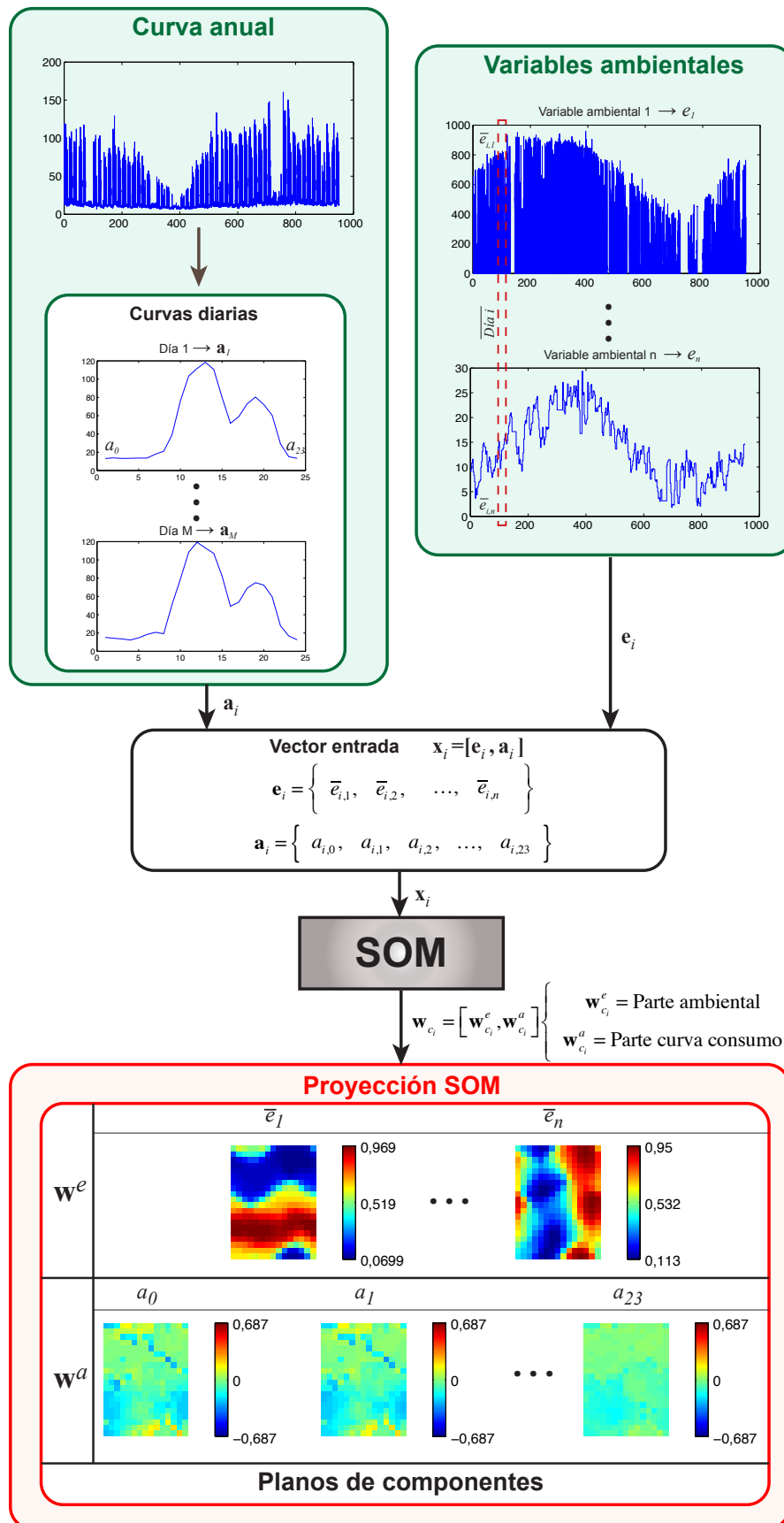


Figura 5.5: Obtención de los modelos utilizando el SOM.

Para poder visualizar estas diferencias entre los modelos de los edificios se utilizan dos mapas de visualización:

- **Mapa estático de módulos de diferencias.** En la retícula de visualización del modelo y_{B_j} , se asigna un color proporcional al valor del módulo de la función diferencia entre modelos en ese punto. Los colores verdosos corresponden a zonas cuya diferencia es cercana a cero, siendo zonas en las que el modelo es similar.
- **Mapas estáticos de componentes de diferencias.** En la retícula de visualización del modelo y_{B_j} se asigna un color proporcional a la componente k -ésima de la función diferencia entre modelos en ese punto. Estos mapas van a determinar las desviaciones de cada una de las variables entre los dos modelos.

A la hora de aplicar este método en edificios, las desviaciones entre cada una de las componentes del modelo, que contienen los consumos para cada hora del día, marcan las diferencias que se producen a lo largo de día entre los modelos de los perfiles. Por tanto, la comparación de estos modelos permite determinar la similitud del comportamiento de consumo energético entre edificios.

Sin embargo, a la hora de comparar los modelos, si consideramos de igual peso todas sus componentes, puede darse el caso de que la elección de la BMU induzca a interpretaciones erróneas al compararse dos neuronas que no representan la misma situación. Ya que el modelo se va a hacer en base a una serie de variables ambientales que permiten relacionar el perfil con estas condiciones, se propone ponderar estas variables ambientales para darles más peso. De esta forma, la comparación de los modelos de los edificios A y B se realiza porque las condiciones ambientales que han generado ese perfil son similares.

5.4. Predicción a largo plazo

Como horizonte de predicción a largo plazo se plantea la necesidad de predecir todo un mes en adelante, fundamentalmente con el objetivo de lograr estimaciones de consumo de dicho mes. De esta forma, se puede tener un conocimiento previo del importe de las facturas y es posible determinar de antemano futuros picos de consumo y anticiparse a dichas situaciones.

La estimación del consumo a largo plazo se basa en la utilización de los modelos generados con el SOM que se obtuvieron en 5.3.1. Los modelos generados con el SOM almacenan información del tipo de día por medio de las variables ambientales. El modelo del SOM contiene una parte que codifica el tipo de día e y otra parte que contiene información de la curva de consumo diaria de la variable a predecir a . Así, el SOM consigue relacionar el modelo que especifica el tipo de día con el modelo de consumo para ese día.

Para calcular la predicción de un día en concreto se calcula la BMU, la cual se determina por medio de una modificación de la ecuación 2.22 (Hollmén y Simula, 1996):

$$c = \arg \min_i \sum_{j \in S} (e_j - w_{ij}^e)^2, \quad (5.14)$$

donde S es el conjunto de subíndices de la parte conocida, que en este caso es la parte de condiciones ambientales. Como salida devuelve la parte que no se ha usado para calcular la BMU con $j \notin S$, es decir los subíndices del vector prototipo \mathbf{w}_c que no se utilizaron en el cálculo de la ganadora. Por lo tanto, la muestra predicha se calcula como $\hat{\mathbf{a}} = \mathbf{w}_c^a$ que se corresponde con las variables del consumo para ese día. Con el fin de mejorar la predicción de la neurona ganadora se suaviza el resultado mediante una red RBF similar a la propuesta por Yin (2002) (ver sección 3.5.1).

La precisión de la predicción mediante estos modelos depende fundamentalmente de la precisión de la estimación de las variables ambientales. Un valor erróneo de estas variables para un determinado día puede hacer que la neurona ganadora no se corresponda con ese tipo de día y provocar una predicción equivocada. Fan *et al.* (2009) demuestra cómo para un caso de predicción de energía eléctrica, el error aumenta conforme varía el error de predicción de la temperatura.

Esta técnica es similar a la propuesta por Mandal *et al.* (2006), en la que se propone una predicción a largo plazo usando días similares, ya que cada neurona del SOM va a guardar los datos de aquellos días idénticos. Martínez Álvarez *et al.* (2011) propone una predicción de potencia similar usando, en este caso, similitudes entre patrones. Se agrupan las curvas de consumo para luego determinar la predicción de consumo en base a la media de las curvas similares. El uso del SOM permite obtener un resultado similar, dado que los vectores prototipo representados por el SOM constituyen medias locales de los datos de entrada.

Dado que lo que se busca con la predicción es obtener una estimación del consumo del edificio y no tanto un valor preciso, para cada día se realiza este proceso con tres estimaciones diferentes de las variables ambientales. La estimación más probable $\hat{\mathbf{e}}_p$, la estimación más desfavorable de las condiciones ambientales $\hat{\mathbf{e}}_d$ y la estimación más favorable $\hat{\mathbf{e}}_f$. La predicción de las variables ambientales se puede obtener por dos vías:

- **Datos de predicción meteorológica.** Estos datos de predicción se pueden obtener a través de varias fuentes de Internet. Para lograr una mayor precisión, se pueden utilizar varias predicciones y calcular una media de todos esos datos con el fin de suavizar la predicción de las variables ambientales. El principal problema de este sistema es que resulta difícil encontrar datos de predicción de la radiación solar, a menos que se haga una estimación a partir de la predicción de nubosidad. El segundo problema es que la mayoría de las predicciones solo son, como máximo, a 15 días, siendo difícil encontrar una predicción a mayor plazo.
- **Modelos meteorológicos.** Se pueden crear modelos meteorológicos para cada día a partir de los datos meteorológicos para ese día en función de los años anteriores. Estos modelos se crean haciendo la media de las variables meteorológicas para ese día con los datos de los años anteriores. Esta predicción meteorológica es poco precisa; sin embargo, permite determinar con el propio modelo desviaciones de las variables ambientales para conocer cuál es la situación meteorológica menos deseable para ese día, lo que permite establecer los peores casos de consumo para esos días.

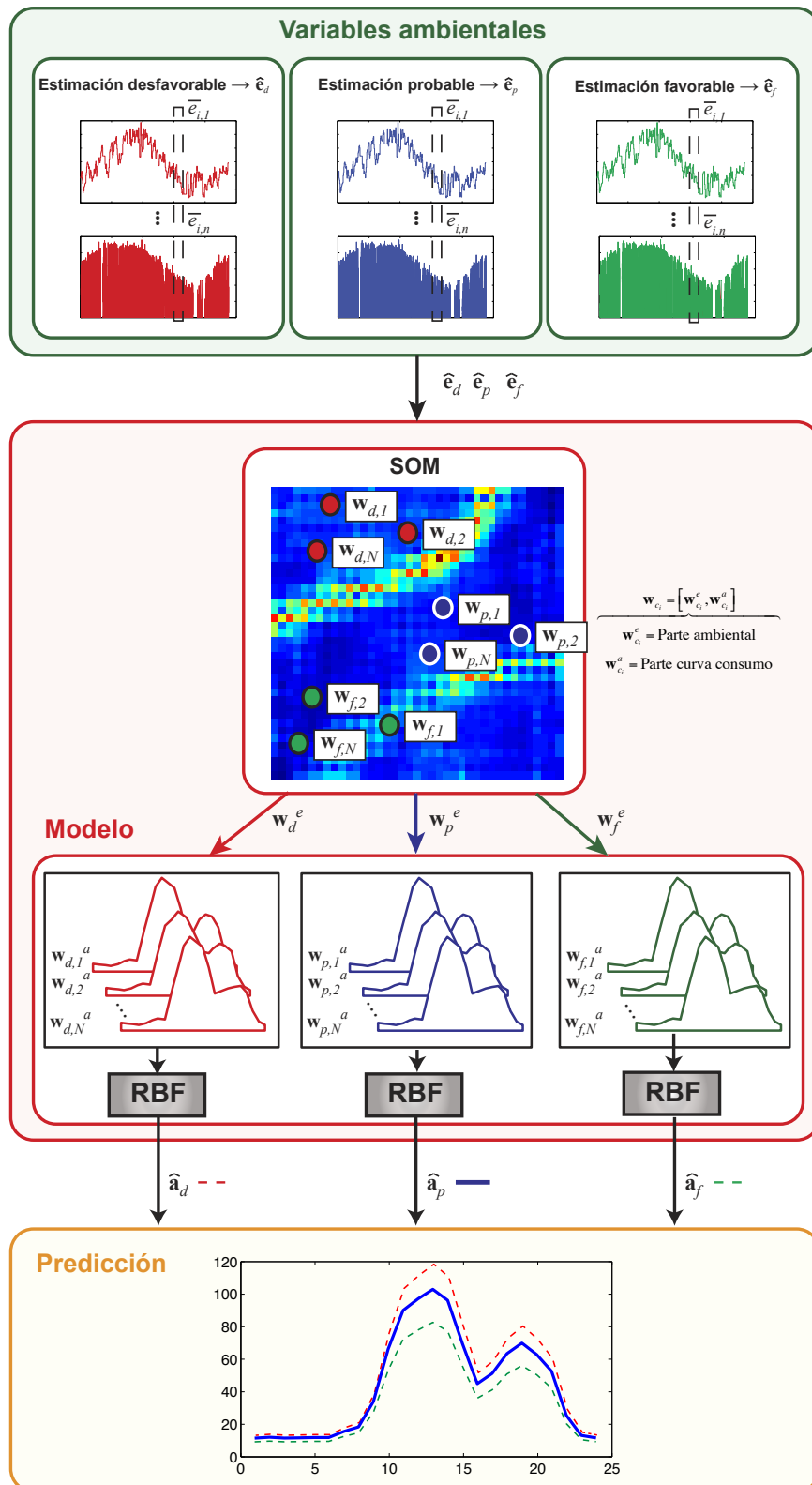


Figura 5.6: Proceso de predicción a largo plazo.

Se obtienen varias predicciones para cada una de las variables, de tal forma que se puedan tener situaciones extremas para cada caso. De esta manera, al predecir utilizando estas desviaciones se puede determinar un posible margen en el cual se va a mover el consumo del edificio. En la figura 5.6 se muestra un esquema del proceso de predicción a largo plazo.

Índices para la medida de la precisión de la predicción

Para determinar la bondad de la estimación es necesario usar un índice que mida lo preciso que es el dato predicho con respecto al valor real del sistema. En Goh y Law (2002) se recogen algunos de los índices más utilizados a lo largo de la bibliografía para determinar lo exacta que es una predicción. Pruebas preliminares llevaron a conclusiones erróneas por una selección incorrecta de indicadores, lo que hizo necesario llevar a cabo un análisis de los distintos índices para ver cual resulta mejor ante el problema de predicción concreto con el que se está trabajando. Los índices utilizados¹ son:

- **Porcentaje del error absoluto medio.** El *Mean Absolute Percentage Error* (MAPE) calcula el error absoluto medido en tanto por ciento. Es un error que depende de la escala de los datos pero que calcula el error expresándolo en tanto por ciento (Armstrong y Collopy, 1992). El principal problema del MAPE es que cuando hay valores nulos en la serie temporal nos encontramos con un término de valor infinito. Por tanto, cuando existen en la serie temporal valores muy pequeños, pueden distorsionar el valor de este índice.

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right|. \quad (5.15)$$

- **Raíz del error cuadrático medio.** El *Root Mean Square Error* (RMSE) es una modificación del índice *Mean Square Error* (MSE) por la cual se calcula la raíz de ese valor (Hyndman y Koehler, 2006). Tiene la ventaja de que no da mayor peso a los errores grandes, por lo que da una mejor idea de la desviación que se ha producido en el cálculo. Por ello suele ser más utilizado a la hora de comparar errores.

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (A_t - F_t)^2}. \quad (5.16)$$

Con el fin de poder comparar valores del RMSE cuando se trabaja con series de diferentes escalas, existen diferentes variantes para normalizar este índice (diferencia entre máximo y mínimo, valor medio, desviación estándar) y dar lugar a lo que se conoce como *Normalized Root Mean Square Error* (NRMSE). A lo largo de esta tesis se va a utilizar el NRMSE normalizado por medio de la diferencia entre el valor máximo y mínimo (Cho *et al.*, 2006):

$$NRMSE = \frac{RMSE}{\max(A_t) - \min(A_t)} \quad (5.17)$$

¹Para todos los índices, A_t se refiere al valor real de las series, F_t recoge el valor predicho por el método a evaluar y N es el número de datos predichos.

Este índice es igual a calcular el RMSE de los datos normalizados y dado que los modelos de predicción se calculan con este tipo de datos, es más sencillo de aplicar.

- **Theil's U.** Este índice también conocido como *Relative Root Mean Square Error* (RelRMSE), es un índice de medida relativo que busca principalmente la comparación entre métodos, más que servir de indicador de la precisión de un método (Theil, 1966). Por ejemplo, para un método de predicción simple ($y(t) = y(t-1)$), se tiene $ThU = 1$. Por lo tanto, para que un método de predicción sea mejor que el método simple, el valor del índice tiene que ser $ThU < 1$, en caso contrario será peor. Dado que las predicciones a lo largo de esta tesis se van a hacer de periodos de 24 horas, el método de predicción simple es $y(t) = y(t-24)$, por lo que el índice queda como sigue:

$$ThU = \sqrt{\frac{\sum_{t=2}^N (A_t - F_t)^2}{\sum_{t=2}^N (A_t - A_{t-24})^2}}. \quad (5.18)$$

5.5. Predicción a corto plazo

La predicción a corto plazo permite una estimación precisa del consumo de un edificio para las siguientes 24 horas. Para este propósito, se comparan cuatro métodos basados en técnicas de reducción de la dimensionalidad y cuantización, viendo así cuál es el mejor método para el caso de los edificios del campus de León. Se utilizan tanto métodos MIMO, que permiten una estimación de múltiples pasos, como métodos directos, que utilizan la información del regresor que se obtendrá por medio de un proceso de selección. Para comparar la eficacia de los métodos, se usan los mismos indicadores de predicción que se usaron para la predicción a largo plazo.

Con el fin de mejorar la predicción, se consideran las variables ambientales que fueron seleccionadas para la creación de los modelos (Gupta y Yamada, 1972). Sin embargo, se modifica un poco la notación y las variables utilizadas. El vector de variables ambientales \mathbf{e} estará formado por las variables ambientales que devuelve la predicción meteorológica para el periodo en el cual se quiere conocer el consumo. La curva de consumos \mathbf{a} pasa a ser en este caso la curva que genera el regresor que se utilizará para predicción:

$$\mathbf{a} = [a(t), \dots, a(t - (m-1))] \quad (5.19)$$

5.5.1. Predicción MIMO

La predicción MIMO, como se explicó en la sección 3.2.2, busca poder predecir varias muestras futuras simultáneamente en base a un regresor. La idea detrás de la predicción MIMO es predecir un día entero adelante en un único paso, por tanto $s = 24$:

$$[\hat{a}(t+s), \dots, \hat{a}(t+1)] = f([a(t), \dots, a(t - (m-1))]) \quad (5.20)$$

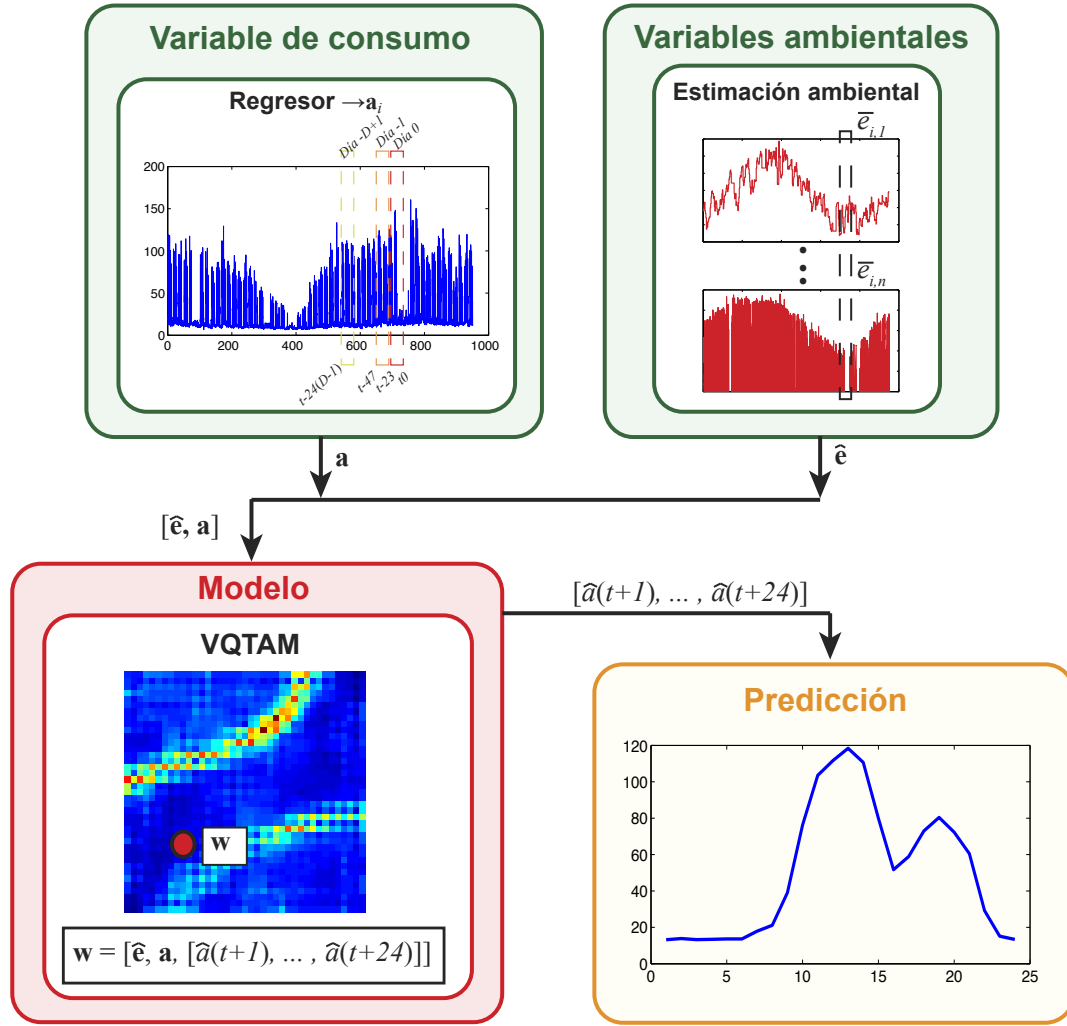


Figura 5.7: Proceso de predicción MIMO con SOM.

Existen pocos métodos que permitan realizar una predicción MIMO (R. Boné, 2002). Sin embargo, tienen como ventaja que el tiempo de cálculo es muy inferior al empleado en caso de una predicción directa. El principal problema de esta técnica es que los errores suelen ser superiores a los de un método directo.

Existen dos técnicas que se pueden utilizar con técnicas de reducción de la dimensión para realizar una predicción MIMO.

Predicción MIMO con SOM

Una de las principales técnicas para la predicción MIMO se basa en el SOM (Vesanto, 1997). Para proporcionar esta funcionalidad, el vector de pesos de la neurona se divide en una parte pasada, que constituye el regresor del modelo, y una parte futura, calculándose la predicción en base a la neurona ganadora con la información de la parte pasada. El método utilizado para predecir es el VQTAM (Barreto *et al.*, 2004), que se describió en la sección 3.5.

A la hora de predecir utilizando el VQTAM, se utiliza un regresor continuo el cual se caracteriza porque contiene todos los instantes de tiempo anteriores al momento a partir del cual se realiza la predicción. Dado que se predice un día entero adelante en un único paso, el tamaño del regresor va a ser de D días de evolución del consumo. A la hora de entrenar el SOM, es imprescindible determinar cuántos días pasados son necesarios para entrenar el algoritmo, con el objetivo de reducir el error de predicción. Para ello, por medio de técnicas de validación cruzada, se selecciona para cada edificio el número de días óptimos.

Con el fin de mejorar este sistema, se incorpora también información de las variables ambientales asociadas a la predicción. De esta forma, el vector de entrenamiento del SOM estará formado por una parte ambiental, que será la media de las variables ambientales estimadas para el periodo en el que se quiere estimar el consumo \mathbf{e} , también contará con el regresor con la información de los consumos de los D días anteriores al periodo a predecir $\mathbf{a}(t)$ y, por último, se compondrá de la parte que codifica la predicción de consumo $\hat{\mathbf{a}}$. El vector del espacio de entrada queda codificado de la siguiente manera, siendo el orden de la serie temporal ascendente:

$$\mathbf{x} = [(e_1, \dots, e_n), (a(t - 24(D - 1)), \dots, a(t)), (\hat{a}(t + 1), \dots, \hat{a}(t + s))]. \quad (5.21)$$

Para que las neuronas se ordenen en base a la información del regresor, se utiliza una máscara para el cálculo de la neurona ganadora durante el entrenamiento. Esta máscara es un vector binario que multiplica, elemento a elemento, el vector de entrada antes del proceso del cálculo de la neurona ganadora, durante el entrenamiento, con el fin de considerar únicamente parte del vector de entrada. La máscara aplicada es:

$$\Omega = [(0, \dots, 0), (1, \dots, 1), (0, \dots, 0)], \quad (5.22)$$

donde solo se considera información aportada por el regresor. Posteriormente, en la predicción, se utilizan tanto las variables ambientales como la parte conocida para calcular la BMU. De esta manera, lo que se logra es que en la etapa de entrenamiento del SOM se ordene en base únicamente a la parte conocida de la serie temporal y se estimen la predicción y las variables ambientales asociadas, confiriendo una mejor ordenación al modelo. En la figura 5.7 se muestra el proceso de predicción utilizando VQTAM.

Predicción MIMO con CCA

Esta técnica, que se presentó en la sección 3.6, consiste en el uso de técnicas de reducción de la dimensionalidad para reducir el número de variables y predecir la trayectoria del espacio de baja dimensión (Chen *et al.*, 2008). De entre las técnicas de reducción de la dimensión se utiliza el CCA porque es la técnica que proporciona mejor continuidad en las proyecciones (Hérault *et al.*, 1999), reduciendo así posibles saltos que se puedan generar en las trayectorias. Dado que estas proyecciones no se van a utilizar para visualización, se puede utilizar la estimación de la dimensión intrínseca (Verveer y Duin, 1995) para generar una proyección mucho más fiable.

Para la utilización de esta técnica, al igual que en las proyecciones que se utilizaban en visualización, se codifica en cada punto la curva de consumo para las siguientes 24 horas,

que será lo que se proyecte, siendo la trayectoria los cambios en la curva de consumo entre periodos de una hora. La predicción será de las siguientes 24 horas, como ocurría en el caso del SOM. Dado que el CCA permite una reconstrucción más sencilla, esta se hace interpolando los k puntos proyectados más cercanos. Estos puntos se ponderan en base a las variables ambientales que se hayan estimado para el periodo de las siguientes 24 horas. En la figura 5.8 se indican los pasos que requiere una predicción utilizando técnicas de reducción de la dimensión, como el CCA.

Predicción MIMO con CCA

1. Separar la variable de consumo en vectores de 24 horas, \mathbf{a} .
2. Proyectar estas curvas \mathbf{A} mediante, CCA \mathbf{Y} .
3. Asignar a cada vector proyectado las variables ambientales correspondientes a ese periodo $[\mathbf{a}_i, \mathbf{e}_i]$.
4. Construir regresores con las trayectorias de la proyección de \mathbf{A} :

$$([y_1(t), y_1(t-1), \dots, y_1(t-n)], [y_2(t), y_2(t-1), \dots, y_2(t-n)]).$$

5. Predecir la trayectoria, calculando cada coordenada por separado:

$$\hat{\mathbf{y}}(t+1) = [\hat{y}_1(t+1), \hat{y}_2(t+1)].$$

6. Reconstruir la curva predicha $\hat{\mathbf{a}}(t+1)$ ponderando los k vecinos de $\hat{\mathbf{y}}(t+1)$ mediante las variables ambientales para esas 24 horas, \mathbf{e} .

La idea es semejante a la aplicada por Lendasse *et al.* (2002), que consiste en predecir la trayectoria proyectada en un espacio de baja dimensión. Esta técnica permite realizar una predicción multidimensional utilizando en este caso solo la predicción de las dos coordenadas de los puntos en el espacio de baja dimensión. El principal problema de la predicción por reducción de la dimensionalidad es que los errores se van acumulando y la cantidad de datos disponible para las proyecciones es demasiado baja, lo que genera proyecciones con elevada distancia entre algunos de los puntos como para que la serie temporal de la trayectoria tenga saltos abruptos y pueda ser predicha con exactitud.

5.5.2. Predicción Directa

La predicción directa, tal y como se explicó en la sección 3.2.1, trata de establecer una función distinta para predecir cada muestra futura de acuerdo a cada función. Por este motivo, un paso importante a la hora de realizar este tipo de predicción es establecer cuál es el mejor tipo de regresor acorde con cada tipo de edificio.

Para ello es necesario hacer primero una selección del regresor con el fin de determinar la dependencia de la salida a predecir con las entradas pasadas. Un análisis de la evolución

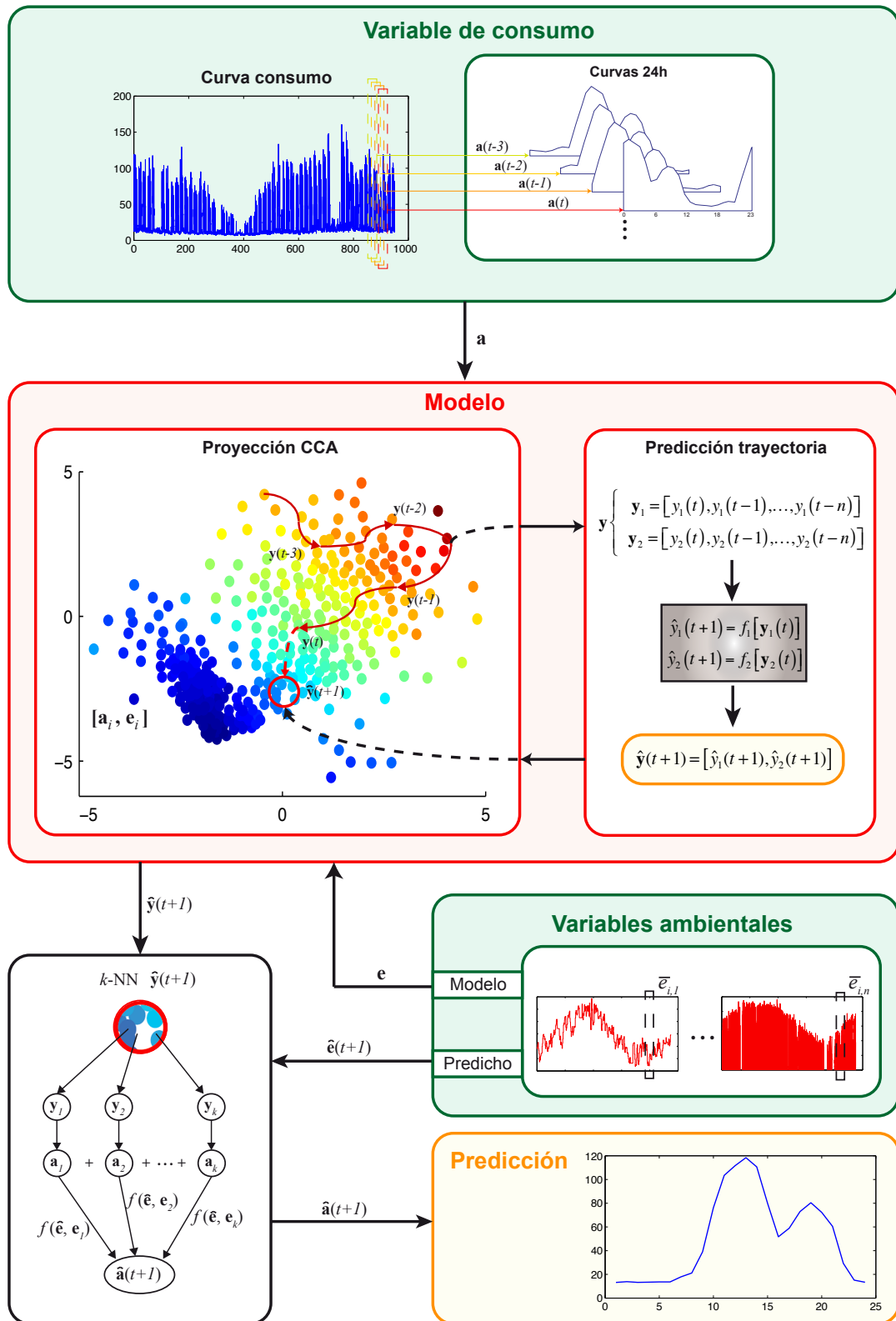


Figura 5.8: Proceso de predicción utilizando CCA.

de la variable a lo largo de las horas permite determinar si la variable es dependiente o no de la hora, con lo que la estrategia pasa de utilizar un regresor secuencial o continuo, en el cual el regresor considera todos los instantes de tiempo anteriores al momento de realizar la predicción, a uno partido, en el que resulta mejor fragmentar el regresor, seleccionando los instantes de tiempo pasados, para realizar la predicción de la variable. En función de la dependencia horaria es mejor analizar únicamente las M_1 muestras anteriores o considerar también información de la misma hora de otros días previos M_2 . El regresor puede ser de tres tipos en función de la parte que se analice:

- **Regresor secuencial.** Es aquel que está compuesto únicamente por información del día previo.

$$\hat{a}(t+s) = f[a(t), a(t-1), \dots, a(t-(M_1-1))]. \quad (5.23)$$

- **Regresor periódico.** También llamado partido, se utiliza información de la misma hora de días anteriores, siendo el número de días considerado M_2 . Este regresor es útil para edificios con una elevada dependencia horaria.

$$\hat{a}(t+s) = f[a(t+s-24), \dots, a(t+s-24M_2)]. \quad (5.24)$$

- **Regresor mixto.** Se utiliza un regresor en el que se analiza la parte secuencial y la parte periódica. Se busca la dependencia no solo del día anterior sino también de la misma hora de días anteriores.

$$\begin{aligned} \hat{a}(t+s) = & f[a(t), a(t-1), \dots, a(t-(M_1-1)), a(t+s-48), \dots \\ & \dots, a(t+s-24M_2)]. \end{aligned} \quad (5.25)$$

En función de la dependencia de la variable energética con la hora del día, se puede determinar cuál es el mejor tipo de regresor para realizar la predicción (Murray *et al.*, 2000). Si la variable de consumo presenta una elevada correlación con la hora del día, se recomienda tener presente la información de la parte periódica (Fay *et al.*, 2003). Por ello, la estrategia de la búsqueda del regresor puede ser de dos formas:

- **Edificios con baja dependencia horaria.** La búsqueda del regresor se limita a su parte secuencial. Para ello, dado que la parte secuencial del regresor requiere de gran cantidad de datos a comprobar, es recomendable que la parte secuencial a analizar tenga un tamaño idéntico a la ventana de predicción (Sorjamaa *et al.*, 2007), esto es, si se van a predecir 24 horas, el tamaño de M_1 será de 24 horas. Para determinar la composición de esta parte secuencial se recurre a métodos de selección con búsqueda de regresor.
- **Edificios con alta dependencia horaria.** La búsqueda del regresor es en dos partes. Por una lado se busca en la parte periódica el tamaño óptimo y una vez que se tiene esta parte se añade información de la parte secuencial como en el caso anterior. El tamaño de M_2 se determina en este caso por un método *wrapper*, ya que el número de días anteriores a la parte secuencial nunca es mayor que 5 y la búsqueda del número óptimo de días se realiza solo añadiendo días.

Una mala selección del regresor genera peores predicciones debido a que, si la entrada es muy grande, aparece el problema de la maldición de la dimensionalidad (Bellman, 1961), aumentando los requerimientos de computación y memoria a la hora de modelizar los datos. Otro problema es que el uso de vectores muy grandes genera modelos complejos difíciles de entender, por lo que modelos simples suelen generar mejores predicciones.

Para la selección del regresor se utilizan dos formas: para la parte periódica se utiliza un método *wrapper* mientras que para la parte secuencial se usa un método de filtrado.

El **método *wrapper*** se basa en entrenar el algoritmo de predicción con todos los datos y comprobar el error del resultado aplicado a una predicción real. Estos métodos, dependiendo del algoritmo de predicción, pueden resultar demasiado costosos en tiempo, ya que algoritmos como el SVR tienen tiempos de entrenamiento muy largos, por lo que entrenar con distintas combinaciones del regresor puede generar tiempos de búsqueda inmensamente grandes (Guyon y Elisseeff, 2003). Son útiles cuando la entrada es pequeña, como es el caso de la parte periódica.

El **método de filtrado** utiliza únicamente la información de los datos de entrada para la selección del regresor. Para poder seleccionar el regresor se utilizan métodos de evaluación que miden la relación entre cada subconjunto de datos sin necesidad de entrenar los algoritmos, por lo que son mucho más eficientes. Como método de filtrado se utiliza un k -NN por ser un método sencillo y por generar resultados óptimos (Sorjamaa *et al.*, 2007).

La aproximación por k -NN (Bishop, 1995) se basa en la idea de que las muestras con entradas similares tienen que tener salidas similares. Para ello, se seleccionan los vecinos más cercanos de acuerdo a una distancia euclídea, calculando la salida correspondiente como el promedio de las salidas de los vecinos más cercanos:

$$\hat{y}_i = \frac{\sum_{j=1}^k y_{j(i)}}{k}, \quad (5.26)$$

donde \hat{y}_j representa la salida estimada, $y_{j(i)}$ es la salida del elemento j vecino más cercano de la muestra \mathbf{x}_i y k indica el número de vecinos seleccionados. Se ha utilizado esta terminología para simplificar la notación. En este caso, \mathbf{x} correspondería a la entrada del regresor, que sería el vector $[a(t), a(t-1), \dots, a(t-(M_1-1))]$, siendo el término escalar y la salida correspondiente $a(t+s)$.

Este método permite seleccionar las variables de entrada, que son aquellas que minimizan el error $y_i - \hat{y}_i$. Solo se requiere como parámetro, para aplicar este método, k , el cual se determina utilizando métodos de validación cruzada. Se opta por el *10-fold* ya que requiere un menor tiempo de ejecución y da buen resultado (Kohavi, 1995).

Dado que la parte secuencial es muy grande, es necesario un proceso de búsqueda distinto al proceso exhaustivo, que consiste en probar todas las posibles combinaciones de las entradas con las salidas. El número de combinaciones posibles es $2^M - 1$, siendo M el tamaño del regresor. Para reducir el tiempo de computación se utilizan dos técnicas que reducen el coste con respecto a una exhaustiva. Existen varias estrategias que reducen el coste que se genera en una búsqueda exhaustiva:

- **Búsqueda hacia delante.** Se parte de un conjunto vacío de entradas B y disponemos de un conjunto F de entradas $A^i, i = 1, 2, \dots, N$, y un conjunto de salidas Y . Se van añadiendo entradas al conjunto B en función del error que genere cada una hasta que el regresor obtenga el tamaño de M variables. En cada paso de este algoritmo de búsqueda se añaden, de una en una, todas las posibles variables de entrada que no están introducidas todavía, y se comprueban los errores que generan de acuerdo al método de filtrado. Una vez que se han probado todas las variables, se selecciona aquella que genera un menor error y se actualiza el regresor. Mediante este método solo $M(M+1)/2$ combinaciones son posibles.
- **Eliminación hacia atrás.** Este caso es el contrario al anterior. Se parte de un conjunto con todas las entradas que se van a considerar, B , y se van eliminando entradas en función del error que genere cada eliminación hasta que el regresor quede vacío. En cada paso se elimina una de las variables y se calcula el error generado al aplicar el k -NN.

Búsqueda hacia delante

1. Inicializar el conjunto B , inicialmente vacío, que va a contener las entradas seleccionadas, y el conjunto F que va a contener el conjunto inicial de entradas de las M posibles.
2. Calcular el conjunto que proporciona el menor error al aplicar el k -NN.

$$A^s = \arg \min_{A^i} k\text{-NN}(\{B, A^i\}, Y) \quad A^i \in F$$

3. Encontrado el mínimo, salvar $k\text{-NN}(\{B, A^i\}, Y)$ y mover A^s de F a B .
4. Repetir los pasos 2 y 3 hasta que el tamaño de B sea M .
5. Comparar los valores de k -NN para todos los tamaños de los conjuntos B y escoger el menor de todos.

Eliminación hacia atrás Es idéntico al anterior salvo que, en este caso, B va a contener inicialmente todas las variables posibles.

Dado que ninguno de los dos métodos garantiza que se obtenga el mínimo óptimo, la solución es utilizar los dos métodos, ya que aun así el número de combinaciones es mucho menor que una búsqueda exhaustiva. Para ello, se ejecutan los dos algoritmos de búsqueda y de los candidatos que devuelva cada uno se escoge aquel que genere un error menor (Thompson, 1978).

Una vez seleccionado el regresor a utilizar, se entrenará el algoritmo de predicción. Para entrenar los modelos existen dos opciones (Sorjamaa y Lendasse, 2006):

- **Utilizar el mismo regresor para predecir todas las horas.** Este caso corresponderá a los regresores periódicos puros, de tal forma que cada uno se creará con

los datos de la hora a predecir pero de los días anteriores:

$$\hat{a}(t+24) = f(a(t), a(t-24), \dots, a(t-24(M_2-1))). \quad (5.27)$$

La ventaja de esta predicción es que solo es necesario generar un único modelo que será útil para todas las horas. El principal problema es que el error de predicción será el mismo para todas las horas, por lo que no se consigue una reducción del error en la predicción de las primeras horas.

- **Utilizar regresores distintos para cada hora del día.** Este caso corresponde a los edificios que no tienen un consumo dependiente de la hora, por lo que se hace necesario crear un regresor distinto en función de la hora a predecir:

$$\begin{aligned} \hat{a}(t+1) &= f(a(t), a(t-1), \dots, a(t-(M-1))), \\ \hat{a}(t+2) &= f(a(t), a(t-1), \dots, a(t-(M-1))), \\ &\vdots \\ \hat{a}(t+24) &= f(a(t), a(t-1), \dots, a(t-(M-1))). \end{aligned} \quad (5.28)$$

Los datos $a(t), a(t-1), \dots, a(t-(M-1))$ se corresponden con los instantes de tiempo devueltos por el algoritmo de selección de regresor y no tienen por qué ser necesariamente los mismos para predecir cada hora. De esta manera cada instante de tiempo a predecir $t+s$ tiene asociado un vector $\mathbf{a}(t+s)$ que recoge los instantes de tiempo devueltos en la selección y, por tanto, la muestra se predice como $\hat{a}(t+s) = f[\mathbf{a}(t+s)]$. La principal ventaja de este método es que se consigue reducir el error para la predicción de las horas cercanas al instante actual, pero se necesita entrenar un modelo para cada hora con el consiguiente coste computacional.

Una vez que se ha seleccionado el regresor, se van a aplicar dos métodos de predicción directa que son el KSOM y un método basado en *clustering*, ambos introducidos en el capítulo 3.

Predicción directa con KSOM

El SOM también es útil para la predicción directa, limitando la parte desconocida de la neurona a un único componente. El principal problema es que se tienen que crear varios modelos, uno por cada regresor, en el caso de que se estén utilizando distintos regresores para cada instante de tiempo.

Dentro de las distintas versiones de algoritmos SOM que existen para predicción, que fueron explicados en la sección 3.5, se hicieron pruebas con las distintas modificaciones que existen de este algoritmo (Prada, 2009). Se utiliza una variante que es el KSOM (Barreto, 2007), el cual se modifica para que las neuronas contengan información no solo de los regresores sino también de las variables ambientales. De esta forma, a la hora de determinar la neurona ganadora, se considera no solo la información de los regresores sino también parte del conocimiento ambiental, al igual que se utilizó en el caso de la predicción MIMO (ver ecuación 5.21), salvo que la parte a predecir estará compuesta de un único componente. En

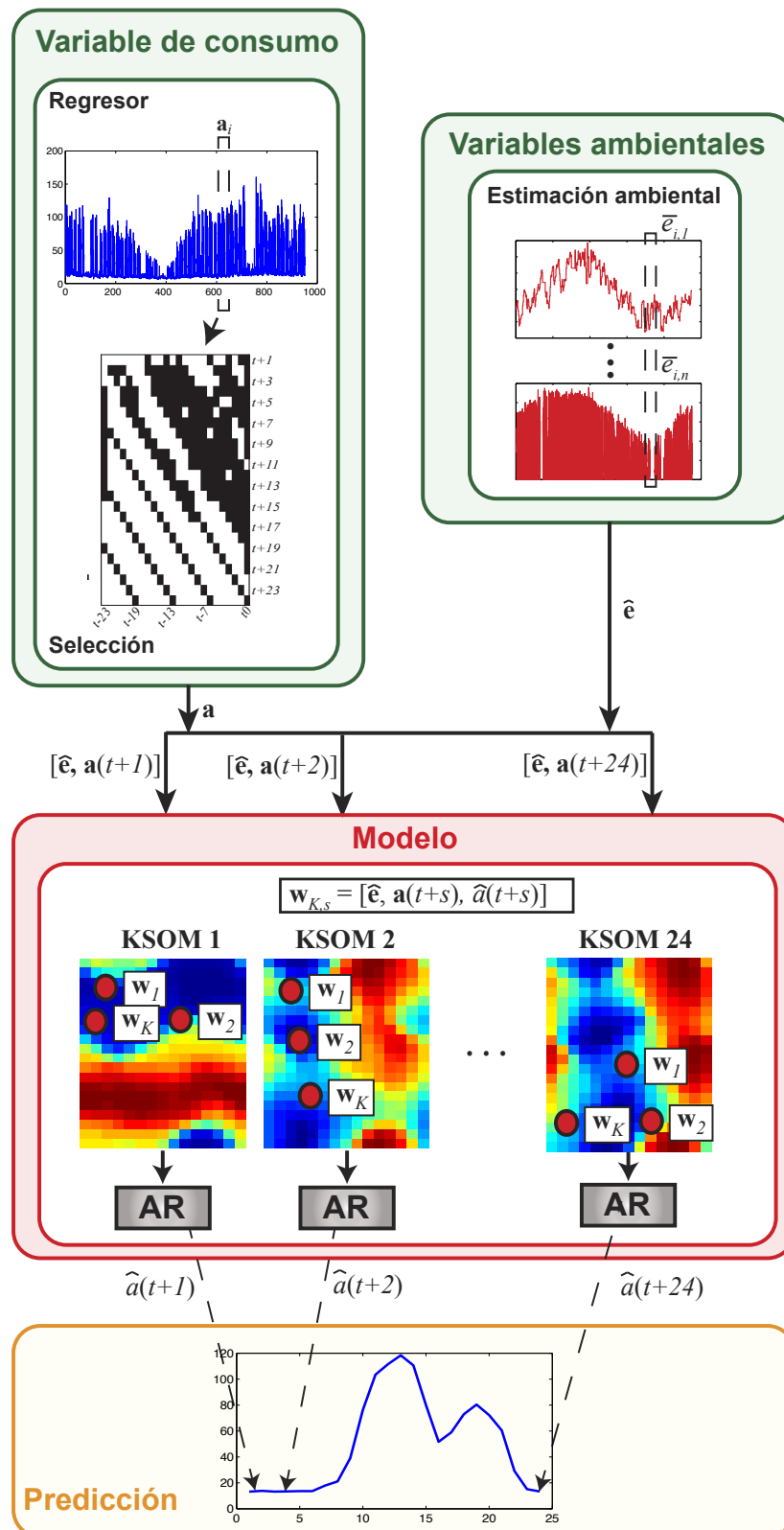


Figura 5.9: Proceso de predicción directa utilizando KSOM.

cuanto a la parte conocida, se genera en base a la información aportada por la selección. En la figura 5.9 se indica el proceso de predicción directa utilizando KSOM. En esta figura el diagrama de selección indica en las filas la composición del regresor para la predicción de cada instante de tiempo, siendo los instantes de tiempo seleccionados los marcados por medio de un cuadro negro.

Predicción directa con *Neural Gas* y SVR

En vez de crear modelos con los regresores y las variables ambientales, otra posibilidad es agrupar los distintos tipos de días que se pueden generar, en base a estas variables ambientales, mediante técnicas de *clustering* y generar un modelo de predicción por cada grupo (Fan y Chen, 2006). En la sección 3.4 se explicaron los distintos métodos de predicción existente y cuáles eran los pasos para combinarlos con una técnica de agrupamiento. De los muchos métodos existentes para regresión de funciones ARIMA, ARMA, SVR (Contreras *et al.*, 2003; Cao, 2003), se ha seleccionado este último porque es el que mejores resultados preliminares generó, a pesar de que es el que resulta más costoso de entrenar.

Predicción directa con *Neural Gas* y SVR

1. Determinar por búsqueda del regresor el regresor óptimo para la predicción.
2. Construir los regresores para cada instante del día **a**.
3. Construir los vectores de variables ambientales para el periodo a predecir **e**.
4. Agrupar los instantes de tiempo en base a las variables ambientales **Q**.
5. Para cada *cluster*, crear un modelo mediante SVR.
 - Un único modelo si el vector es periódico.
 - Veinticuatro modelos (uno para cada hora) si el vector no es periódico.

La idea fundamental es realizar predicciones a un día, es decir, en cualquier hora del día se van a predecir las siguientes 24 horas. Para ello, las condiciones ambientales a considerar son la media de las variables para ese periodo. El algoritmo se compone de dos fases:

- **Clustering.** Se agrupan los distintos tipos de periodos de 24 horas en función de las variables ambientales **e** para ese tramo. Existen variantes que utilizan como método de agrupamiento de las variables ambientales el SOM (Ismail *et al.*, 2011); sin embargo, dado que en esta primera fase se busca únicamente obtener los grupos de días, resulta más eficiente usar un algoritmo específico de *clustering*, por lo que se recurre a uno que constituye una modificación del SOM, que es el *Neural Gas*, y que proporciona buenos resultados (Martinetz *et al.*, 1993).
- **Regresor SVR.** Una vez que se han obtenido los grupos **Q** con el *clustering*, para cada grupo se asigna un modelo de regresión calculado por medio de SVR. En función

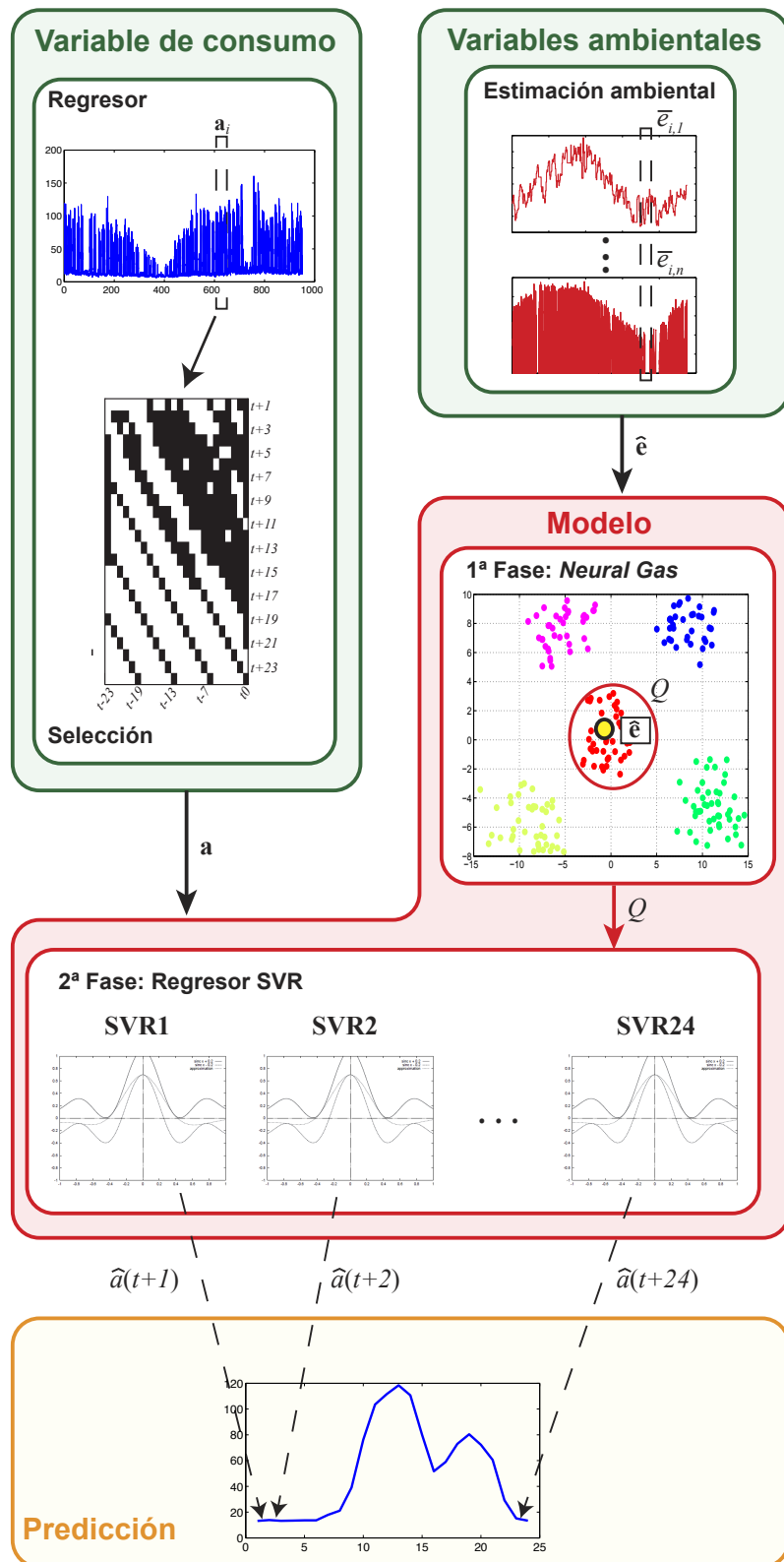


Figura 5.10: Proceso de predicción directa utilizando *Neural Gas* y SVR

de que el regresor sea periódico o no, será necesario calcular un único modelo para todo el periodo o calcular un modelo para cada instante de tiempo (Fan *et al.*, 2009). Para entrenar los modelos no-lineales se utilizarán las *Least Squares Support Vector Machines*, que solucionan el problema de mínimos locales (Suykens *et al.*, 2002a).

En la figura 5.10 se muestran los pasos para la creación del modelo de predicción.

Experimentación y resultados

Los experimentos que se van a realizar son fundamentalmente cuatro. El primero de ellos tiene como objetivo analizar el comportamiento de los perfiles energéticos de consumo, de tal forma que se pueda determinar la dependencia del perfil de consumo con determinadas variables ambientales y relacionar estas con un nuevo índice de actividad que dé idea del uso del edificio. Esta variable resulta útil en especial para los edificios cuyo consumo depende fundamentalmente de la docencia.

El segundo de los experimentos sirve para comparar los edificios entre sí mediante dos métodos diferentes. Por un lado, el uso de mapas de diferencias, que se han probado con éxito para procesos industriales, como se refirió en el capítulo anterior, en este caso va a permitir estudiar las desviaciones que se producen, analizando de forma visual, qué edificios presentan comportamientos similares. Por otro lado, se van a comparar los perfiles proyectados mediante técnicas de proyección. La forma que adquieran los datos proyectados es la que marcará la similitud en el comportamiento de los edificios.

El tercer experimento consistirá en comparar distintos métodos de predicción utilizando los modelos y la información obtenida en el primer experimento. En función de estos datos se seleccionará un método distinto para realizar la predicción o se tendrán en cuenta distintos regresores con el fin de poder realizar la predicción de los perfiles.

El cuarto de los experimentos consistirá en proyectar los datos de las facturas con el fin de poder obtener información de estas y poder comparar los datos que almacenan.

De los 32 edificios de los que dispone la Universidad, enumerados en la tabla 4.1, hay que descartar los edificios de Educación, Sucursal Bancaria, Complejo La Serna, Enfermería y Cafetería III, bien porque los medidores no funcionan correctamente o bien porque no han tenido consumo, por lo que no se pueden tener en consideración para el análisis.

Para la ejecución de los algoritmos se han utilizado las siguientes librerías:

- *SOM Toolbox* de la Universidad de Helsinki¹ (Vesanto *et al.*, 1999).
- *SOM Toolbox* de la Universidad de Oviedo.
- *Toolbox* para la reducción de la dimension² (Van der Maaten *et al.*, 2007).

¹<http://www.cis.hut.fi/projects/soomtoolbox>

²http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html

- *Toolbox* para LSSVM³ (Suykens *et al.*, 2002b).

Para la generación de las visualizaciones se ha utilizado *Processing* (Reas y Fry, 2007).

6.1. Obtención del espacio de entrada

Como paso previo a la generación de las proyecciones y a los modelos, es necesario determinar las variables que componen los vectores del espacio de entrada. Estos vectores están compuestos por las curvas diarias de la variable de consumo energético a analizar y por una serie de variables ambientales que estarán relacionadas con el consumo y la actividad del edificio. Algunas de estas variables se seleccionan por medio de un análisis de correlaciones (variables analógicas), mientras que otras variables que están más relacionadas con la actividad requieren de un mayor procesamiento y análisis (variables binarias).

6.1.1. Selección de la variable de consumo

La variable energética seleccionada para los experimentos es la potencia activa eléctrica. La curva diaria se genera con los valores medios de la potencia para cada hora, por lo tanto los vectores que componen las curvas de consumo son:

$$\mathbf{a} = \{ P_0, P_1, P_2, \dots, P_{23} \}, \quad (6.1)$$

donde P_h es la potencia media para la hora h de ese día.

La variable de consumo se somete a un análisis de correlación horaria con el fin de conocer la dependencia entre la potencia consumida por los edificios y la hora del día. Cuanto mayor sea esa dependencia, más regular será el comportamiento de consumo del edificio. Este análisis se realiza aplicando la ecuación 5.4 y mostrando los resultados por medio de representaciones de curvas de nivel, como los de las figuras 6.1 y 6.2.

En la figura 6.1 se muestra la correlación horaria para el edificio de Filosofía, que es el que se ha considerado más dependiente de todos, y en la figura 6.2 se muestra la correlación horaria para el Animalario, que se considera el menos dependiente. Para determinar esta dependencia se calcula la correlación entre los datos para cada hora del día. Cuanto mayor sea el paralelismo de sus curvas de nivel con la diagonal de la gráfica menor será el grado de dependencia de la potencia de la hora. En las figuras se muestran como referencia dos líneas discontinuas, de modo que cuanto mayor paralelismo exista entre las curvas de nivel con estas líneas, menor será la correlación horaria.

En la tabla de correlaciones 6.1 se muestran los valores mínimos de correlación entre horas que se dan en cada edificio. Cuanto más elevado es el valor mínimo de la correlación menor será la dependencia del consumo de la potencia con la hora del día. Si el valor mínimo es muy bajo puede darse el caso de que exista alguna hora que tenga un consumo diferente a las del resto sin ser dependiente de la hora, por lo que en esos casos se requiere ver el gráfico de curvas de nivel. Así, Filosofía o Ciencias del Trabajo tienen un valor bajo y el

³<http://www.esat.kuleuven.ac.be/sista/lssvmlab/>

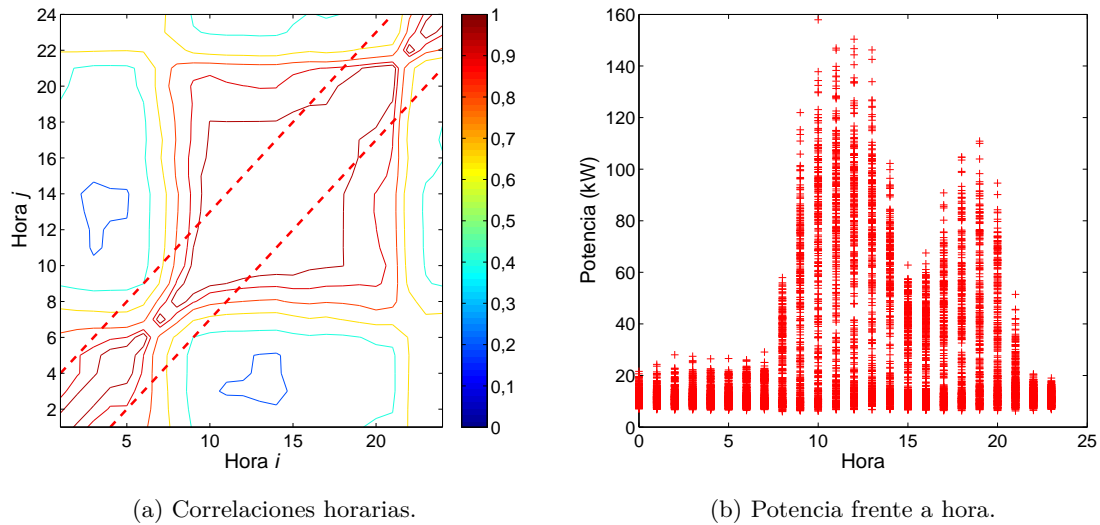


Figura 6.1: Análisis de la dependencia de la potencia con la hora en el edificio de Filosofía.

análisis de las curvas muestra una dependencia de la hora, mientras que Animalario o Clínica Veterinaria presentan valores altos.

6.1.2. Selección de las variables ambientales

A la hora de analizar las curvas de consumo, es interesante conocer la dependencia de los perfiles de consumo con las variables ambientales, con el fin de seleccionar aquellas que tienen una gran relación con el consumo. Dentro de las variables ambientales de las que se dispone, estas pueden estar relacionadas con condiciones meteorológicas, actividad del edificio o instante de tiempo.

De entre las variables meteorológicas leídas por la estación, están la humedad, la temperatura y la radiación solar. Se calcula la correlación entre la potencia media de cada día y el valor medio de la variable ambiental, para estudiar la dependencia entre unas variables y otras.

Se muestra la correlación para los edificios de Filosofía y Animalario para ver la dependencia en el caso de un edificio que tiene una elevada correlación horaria y otro que no la tiene. En la figura 6.3 se puede ver la gráfica de las potencias medias frente a cada una de las variables ambientales para el edificio de Filosofía. En la figura 6.4 se pueden observar las correlaciones en el caso de Animalario.

En la tabla 6.1 se recogen todas las correlaciones entre los edificios y estas variables. De esta tabla y de las figuras, se puede deducir cómo influyen las distintas variables meteorológicas en el consumo. Existen edificios que no tienen correlación alguna, como es el caso de Animalario, mientras que otros edificios como Filosofía poseen una correlación elevada con la temperatura y la radiación de forma inversa, esto es, a menor temperatura mayor es el consumo de potencia y lo mismo ocurre con la radiación. Como era previsible,

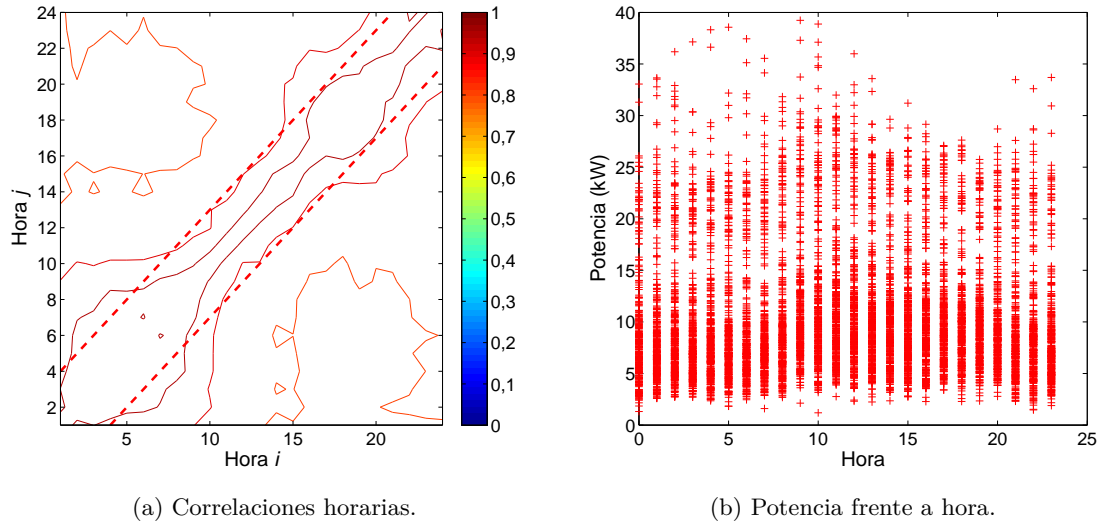


Figura 6.2: Análisis de la dependencia de la potencia con la hora en el edificio de Animalario.

esto conlleva una conclusión lógica, que es que cuanto más frío o menor radiación, más se utilizan las luces y calefacciones.

De esta tabla también se puede deducir que, por lo general, la humedad tiene poca correlación con la potencia en comparación con las otras dos variables. Si a esto le añadimos que la correlación existente entre la humedad y la temperatura es $-0,443$, siendo este en valor absoluto igual o inferior al de la correlación humedad-potencia para muchos edificios, se puede desechar esta variable ambiental para análisis posteriores.

Las variables relacionadas con el instante de tiempo, que son incluidas en el análisis, son el día de la semana y la hora. Estas variables se caracterizan por las discontinuidades que presentan (23 p.m - 0 a.m.) y porque generan una elevada distancia, que modifica el resultados de las proyecciones. Por ello, estas variables se descomponen como dos coordenadas, x e y , que se corresponden con las manecillas de un reloj y evitan esas discontinuidades. Estas variables son interesantes para considerar la estacionalidad del consumo. La hora tendrá su utilidad en el caso de la predicción; sin embargo, para el caso de las proyecciones no se utiliza, dado que en cada vector se incorpora toda la curva diaria, por lo que esta variable va implícita en el vector. En cuanto al día de la semana, que es la otra variable a ser considerada, es incorporada en el cálculo del índice de actividad, ya que mejora los resultados de la generación de este índice.

Índice de actividad

El índice de actividad es la variable que agrupa todas aquellas variables ambientales que están relacionadas con el uso del edificio y que, por lo general, son binarias. Para determinar este índice de actividad se agrupan los días en función de una serie de variables, consignadas en la tabla 6.2. Esto va a permitir agrupar los distintos tipos de días que se pueden dar en el calendario en función de las variables que los definan, convirtiendo varias variables binarias en una única variable continua.

ID	Edificio	Corr T-P	Corr Hum-P	Corr RS-P	min Corr h
E1	CRAI-TIC	0,525	-0,258	0,453	0,564
E2	Radio	-0,185	-0,049	0,047	0,125
E3	Industriales I	-0,521	0,262	-0,313	0,337
E4	Aulario	-0,670	0,477	-0,526	0,294
E5	Industriales II	-0,033	0,044	0,051	0,455
E6	Filosofía	-0,713	0,528	-0,648	0,181
E7	Derecho	-0,686	0,413	-0,570	0,256
E9	Ciencias Trabajo	-0,772	0,536	-0,621	0,027
E10	Clínica Veterinaria	-0,303	0,152	-0,204	0,619
E11	Veterinaria	-0,536	0,214	-0,213	0,412
E12	Cafetería I	-0,681	0,378	-0,523	0,163
E13	Biología	-0,651	0,424	-0,523	0,331
E14	Servicios	-0,703	0,409	-0,519	0,279
E15	Desarrollo Ganadero	-0,315	0,181	-0,235	0,578
E16	Animalario	-0,306	0,109	-0,241	0,808
E17	INCAFD	-0,827	0,551	-0,714	0,373
E18	Pabellón Deportivo	-0,678	0,364	-0,534	0,593
E19	Frontón	-0,196	0,118	-0,101	-0,347
E20	Biblioteca Central	0,004	-0,097	0,211	0,019
E22	Cafetería II	-0,433	0,208	-0,322	0,524
E23	Molecular	0,208	-0,065	0,093	0,667
E24	Complejo Agrícolas	-0,483	0,332	-0,444	0,440
E25	Colegio Mayor	-0,618	0,412	-0,352	0,537
E26	Complejo Rectorado	-0,591	0,412	-0,433	0,214
E28	Minas	-0,731	0,526	-0,609	0,352
E29	Centro Idiomas	-0,739	0,382	-0,535	0,259
E31	Económicas	-0,613	0,386	-0,486	0,244

Tabla 6.1: Correlaciones para los edificios analizados para temperatura (T), humedad (Hum), radiación solar (RS) y mínimo de la correlación horaria (min Corr h).

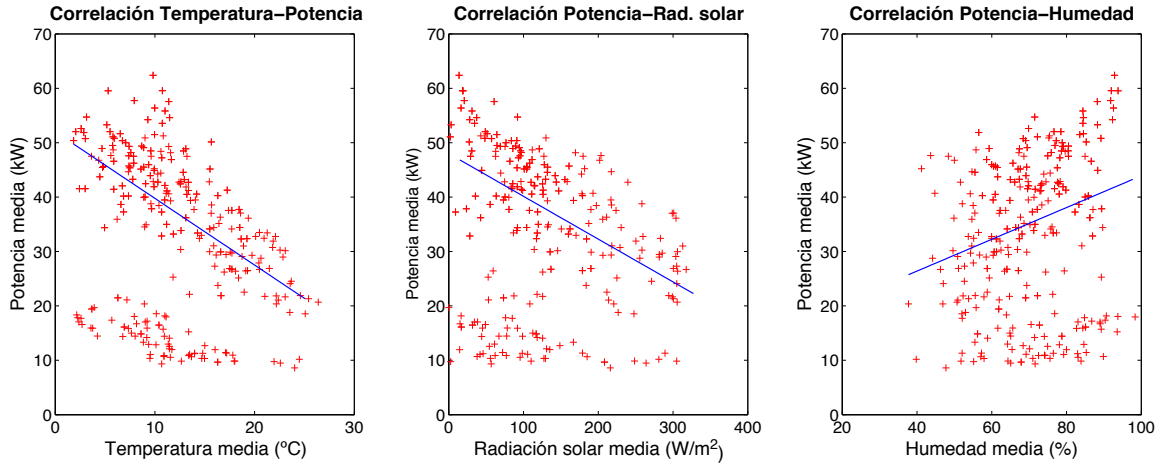


Figura 6.3: Correlaciones entre las variables meteorológicas y la potencia media diaria consumida para el edificio de Filosofía.

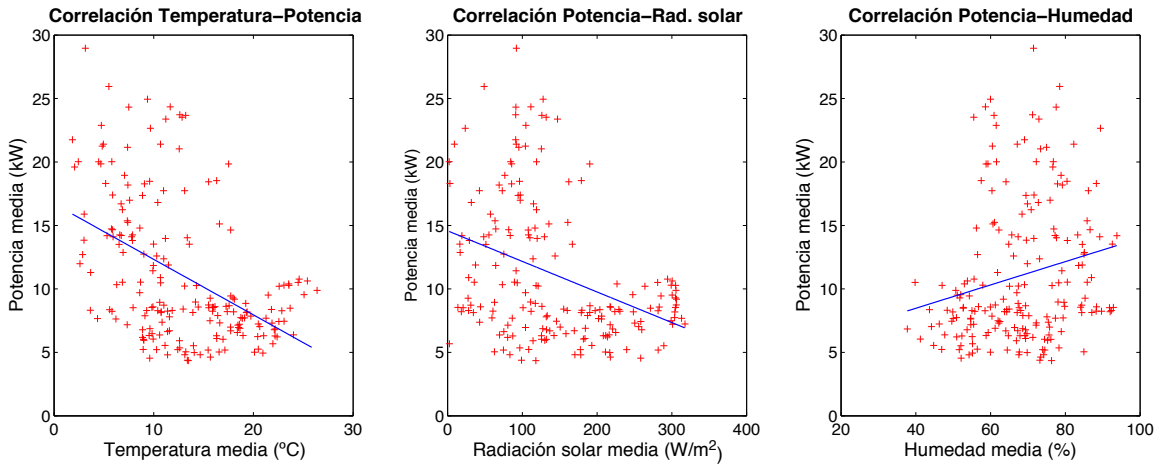


Figura 6.4: Correlaciones entre las variables meteorológicas y la potencia media diaria consumida para el edificio de Animalario.

De los métodos explicados en la sección 2.4, se prueban todos con distinto número de *clusters*, con el fin de establecer el mejor método y el número óptimo de particiones. La figura 6.5 muestra los resultados del *clustering* de las variables que componen el índice de actividad, para los diferentes métodos. Para la evaluación del agrupamiento se ha utilizado el índice de *Davies-Bouldin*, ya que es uno de los más empleados en la bibliografía. Este índice no solo va a indicar cuál de los métodos es el más adecuado sino también el número de particiones necesarias.

Analizando la figura 6.5, se puede observar que no existe un método de *clustering* que destaque sobre los demás. A la vista de los resultados, no se aprecia ningún mínimo que determine la elección óptima del número de *clusters* a utilizar. Por ello, para evitar una solución que implique un *cluster* por dato, se usa una partición de datos que muestre un mínimo local en el vértice de la curva, por lo que el número de particiones queda fijado en torno a 24. Dado que para este valor el *Neural Gas* es el que menor índice genera, es el método utilizado para la partición de los datos.

Identificador	Variable	Tipo de variable
D_x	Día de la semana. Coordenada x	Senoidal
D_y	Día de la semana. Coordenada y	Senoidal
H	No lectivo (Vacaciones Verano)	Binario
E	Exámenes	Binario
F	No lectivo (Festivo)	Binario
X	No definido	Binario

Tabla 6.2: Variables para el cálculo del índice de actividad.

Una vez determinados los distintos tipos de días que existen en el calendario, es necesario asignarles un valor que dé idea del nivel de actividad del edificio y que permita convertir esta variable en una variable continua. Para ello, una vez obtenidos los distintos tipos de día, a cada tipo se le va asignar un valor en función de la potencia media consumida para todos esos tipos de día.

En principio se pueden utilizar dos tipos de índice de actividad:

- **Índice de actividad común a todo el campus.** Es un único índice que representa la actividad del campus. Esta variable está basada en la cantidad de clases que está previsto impartir en toda la universidad para ese día. Ya que todos los centros tienen un nivel de actividad similar en lo que a clases se refiere, el cálculo de un índice de actividad común permite reducir la correlación que se genera al utilizar la potencia para calcular esta variable. El cálculo de esta variable se hace con los datos de potencia de todo el campus, puesto que la mayor parte de la actividad del campus es educativa y es la que más influye en el consumo.
- **Índice de actividad independiente para cada edificio.** Es una variable por edificio, complicando bastante el cálculo de la variable y aumentando el tiempo de procesado. El problema de usar este índice es que en edificios que no tienen un consumo periódico esta variable no aportaría ninguna información, puesto que el consumo no depende del tipo de día, lo que complica la obtención de modelos y la posterior comparación entre datos. Además, a esto hay que añadir el hecho de que los edificios con baja dependencia horaria tienen, en algunos casos, consumos idénticos para todos los días, con lo que el valor de este índice es siempre el mismo.

La razón para utilizar un índice de actividad continuo es que generalmente los algoritmos de proyección y modelado no suelen funcionar bien con variables binarias, complicando la visualización. Si a esto le añadimos que se encuentran varias variables binarias los resultados pueden no ser satisfactorios, por lo que un agrupamiento previo reduce el problema. Además, en el proceso posterior de modelado, el uso de este índice continuo va a permitir a un experto la modificación de su valor, con el fin de poder visualizar la evolución de las curvas de consumo asociadas.

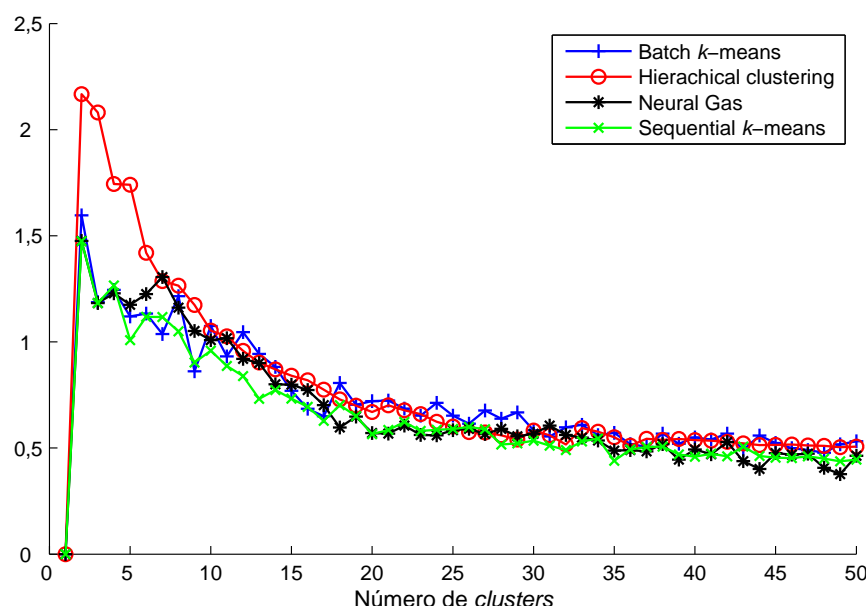


Figura 6.5: Índice Davies-Bouldin para los distintos métodos de *clustering*.

Por lo tanto, una vez que se ha realizado el análisis de las variables ambientales, estas se unen a las curvas de consumo para generar los vectores que componen el espacio de entrada. El espacio de entrada está compuesto por datos de 27 edificios durante un año, por lo que el número de muestras es de 9855. En cuanto a la dimensión de este espacio de entrada, es de 24 variables de consumo y 3 variables ambientales. En la tabla 6.3 se recogen las variables que componen los vectores de entrada.

6.2. Proyección de los perfiles de consumo

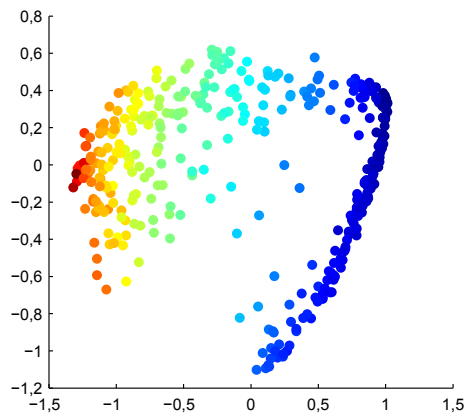
Para generar las proyecciones que permiten analizar tanto las curvas de consumo como su dependencia con las variables ambientales, se proyectan los vectores generados en el apartado anterior. La proyección de las curvas implica primero la selección de la proyección óptima para este tipo de datos. Una vez seleccionada, se hará una proyección de cada edificio, por separado, para ver, en cada caso, la forma que adquiere la proyección de las curvas y poder estudiar la dependencia de cada curva con las variables ambientales. Por último, se proyectan todos los edificios con el fin de poder compararlos entre sí.

6.2.1. Selección de la proyección

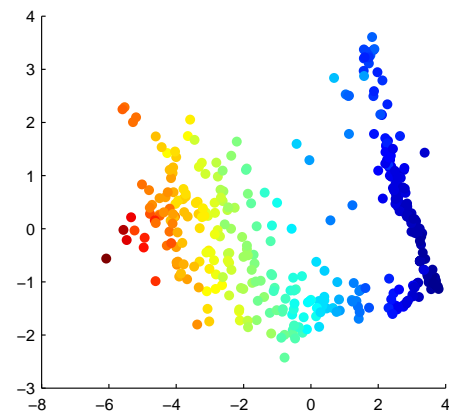
Para la creación de las proyecciones en el espacio de baja dimensión (espacio de visualización) de los perfiles diarios, se prueban todas las técnicas que fueron descritas en el capítulo 2. Para realizar las proyecciones se tendrán en cuenta únicamente los datos de potencia de un año.

	Etiqueta	Descripción	Unidades
Ambiente	I_{ac}	Índice de actividad	-
	T	Temperatura	°C
	R	Radiación solar	W/m ²
Perfil	P_0	Potencia media hora 0	kW
	P_1	Potencia media hora 1	kW
	P_2	Potencia media hora 2	kW
	P_3	Potencia media hora 3	kW
	P_4	Potencia media hora 4	kW
	P_5	Potencia media hora 5	kW
	P_6	Potencia media hora 6	kW
	P_7	Potencia media hora 7	kW
	P_8	Potencia media hora 8	kW
	P_9	Potencia media hora 9	kW
	P_{10}	Potencia media hora 10	kW
	P_{11}	Potencia media hora 11	kW
	P_{12}	Potencia media hora 12	kW
	P_{13}	Potencia media hora 13	kW
	P_{14}	Potencia media hora 14	kW
	P_{15}	Potencia media hora 15	kW
	P_{16}	Potencia media hora 16	kW
	P_{17}	Potencia media hora 17	kW
	P_{18}	Potencia media hora 18	kW
	P_{19}	Potencia media hora 19	kW
	P_{20}	Potencia media hora 20	kW
	P_{21}	Potencia media hora 21	kW
	P_{22}	Potencia media hora 22	kW
	P_{23}	Potencia media hora 23	kW

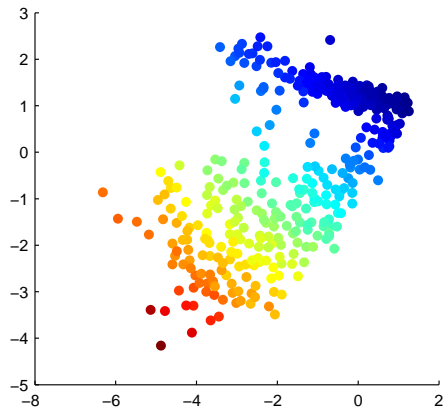
Tabla 6.3: Variables del espacio de entrada.



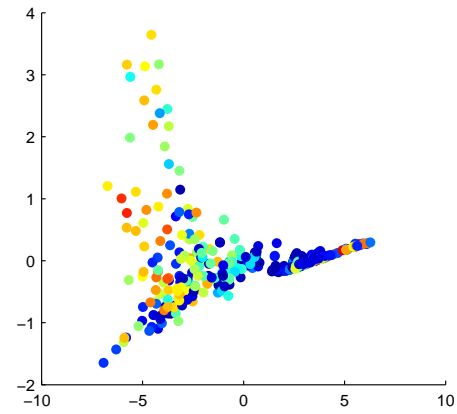
(a) MDS.



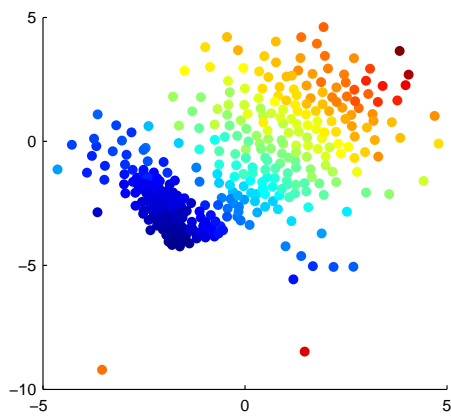
(b) Isomap.



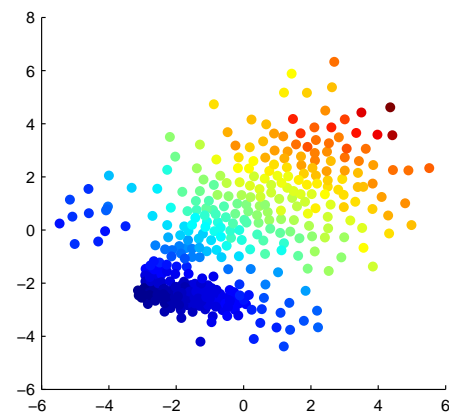
(c) Sammon.



(d) GNLM.



(e) CCA.



(f) CDA.

Figura 6.6: Comparación de los métodos de reducción de la dimensionalidad para los perfiles eléctricos diarios (I).

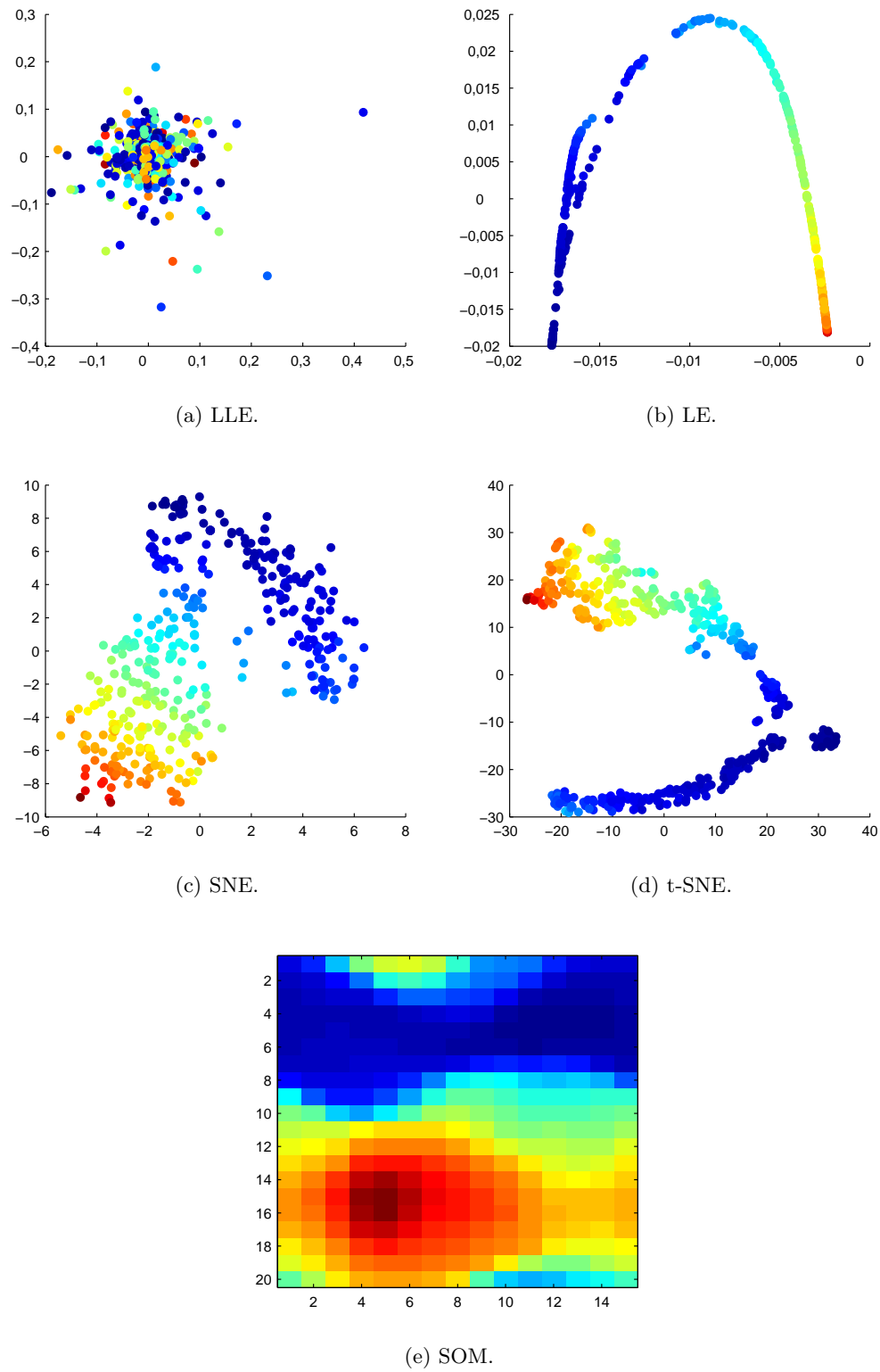


Figura 6.7: Comparación de los métodos de reducción de la dimensionalidad para los perfiles eléctricos diarios (II).

Las variables ambientales que se utilizan son la temperatura, radiación solar y el índice de actividad.

Las figuras 6.6 y 6.7 muestran los resultados de aplicar las proyecciones para el edificio de Filosofía, que es el que presenta el consumo más dependiente de la hora del día de todos los edificios que componen el campus, como se puede ver en la tabla 6.1. El motivo de utilizar este edificio es comprobar que la proyección seleccionada es útil para observar esta dependencia. En estas gráficas, el color del punto depende de la potencia media consumida en ese día, de tal forma que valores rojos indican un consumo elevado y los valores azules indican un consumo bajo.

A la vista de estos mapas se puede observar como el LLE y el GNLM no consiguen encontrar una solución satisfactoria y generan proyecciones en las que todos los puntos aparecen mezclados. Estos dos métodos no serán tenidos en consideración para análisis posteriores. En cuanto al SOM, dado que genera un mapa que conserva únicamente la topología y realiza una cuantización de los datos, no resulta útil para ver la proyección de todos los puntos y se utilizará más adelante con otro propósito.

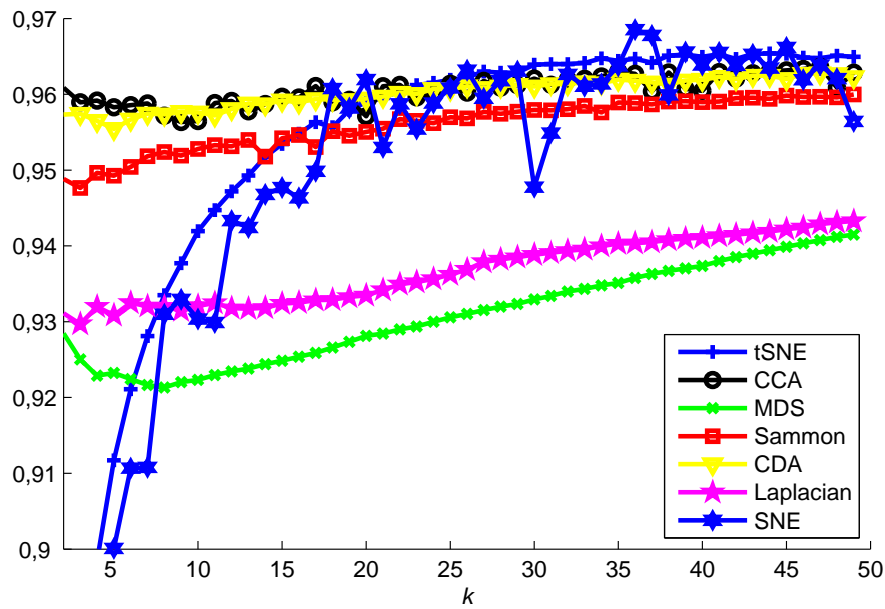
Las proyecciones consiguen distribuir de una forma lógica los datos agrupando aquellos días que presentan un consumo de potencia reducido, mientras que los puntos de elevado consumo están lejos de estos. Más adelante, con la herramienta de *Processing* para la comparación de edificios, se puede ver que en el caso de edificios con una elevada dependencia horaria los perfiles se ordenan siguiendo dos direcciones principales (ver figura 6.15).

- La primera dirección es la que indica la potencia consumida a lo largo del día, estando en un extremo los de bajo consumo y en el opuesto los de consumo elevado.
- La segunda dirección, perpendicular a la primera, indica la distribución de los picos en el perfil para consumos similares.

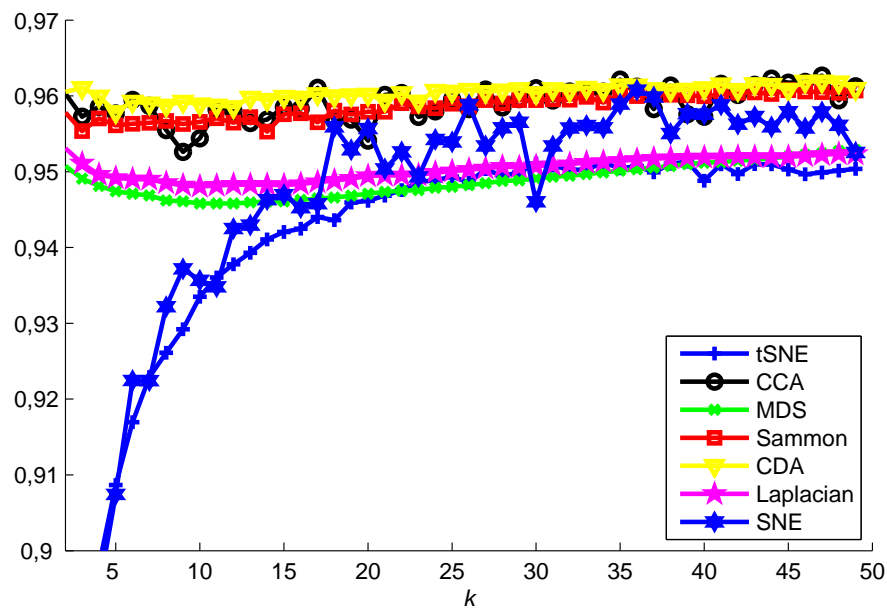
Este perfil generalmente varía en función de la hora en la que se produce el pico de consumo. Métodos como el t-SNE y el LE tienden a agrupar más los datos en una dirección que adquiere una forma más o menos curva, mientras que el resto de métodos distribuye más los datos.

Con todos los métodos, menos el LLE y el GNLM, se van a calcular los parámetros de confianza y continuidad explicados en la sección 5.2.1 y con el fin de completar la comparación de las visualizaciones y ayudar a escoger el método de reducción de la dimensionalidad más apropiado. Dado que los métodos tienen una inicialización aleatoria, el resultado varía con cada ejecución de los algoritmos. Por ello, para el cálculo de estos índices, el algoritmo se ejecuta varias veces para cada valor de k y se utiliza la media entre los valores obtenidos como valor para el índice⁴. En el caso de métodos que tengan como parámetro la vecindad, para cada valor de k se ejecutarán varias veces los algoritmos con distintos valores de vecindad, tomándose igualmente la media de los valores.

⁴Hay que tener presente que este valor de k se refiere al número de vecinos a considerar para determinar la continuidad y fiabilidad del método y que no coincide con el valor de vecindad necesario como parámetro para algunas de las técnicas de reducción de la dimensionalidad.



(a) Fiabilidad.



(b) Continuidad.

Figura 6.8: Índices de fiabilidad y continuidad para los métodos de reducción de la dimensionalidad.

En la figura 6.8a se muestran los valores que toma el índice de fiabilidad en función del valor de k . Se puede ver cómo los mejores valores de fiabilidad se dan en el CCA para valores bajos de k , mientras que para valores altos resultan mejores técnicas de proyección el SNE y el t-SNE. El hecho de que estas dos últimas técnicas sean mejores para grandes valores de k demuestra que funcionan mejor a la hora de conservar la topología de todos los datos en general, mientras que para vecindades pequeñas no logran buenos resultados. Además, el t-SNE, al utilizar una distribución t-Student, que da más peso a puntos lejanos, proporciona mejores resultados que el SNE para valores elevados de k . Los métodos que han dado peores resultados son el LE y el MDS, mientras que el Isomap ni siquiera se muestra en las gráficas porque el índice es menor de 0,9. Esto se debe a que el número de puntos de salida nunca es igual a los puntos de entrada, por lo que el índice se dispara al existir puntos en el espacio de entrada que no existen en el de salida.

Para el caso de la continuidad (ver figura 6.8b), los mejores métodos son el CDA y el CCA, mientras que el t-SNE ni siquiera se acerca para valores altos de k , aunque el SNE casi consigue igualar los resultados del CDA para valores altos de k . Ya en las figuras 6.6 y 6.7 se podía ver como el CDA tiende a distribuir mucho los datos en el espacio de salida, lo que concuerda con valores altos de continuidad. El SNE y el t-SNE, sobre todo este último, tienden a agrupar más los datos, lo que resulta en una peor continuidad.

A tenor de estos resultados, junto con el análisis visual de las proyecciones, los métodos escogidos para realizar las proyecciones de los perfiles diarios van a ser el SNE y el t-SNE, ya que el número de puntos con el que se va a trabajar va a ser elevado, sobre todo cuando se quieran proyectar los datos de todos los edificios. La elección de estos dos métodos se basa fundamentalmente en que las proyecciones que generan agrupan mejor los datos y que el grado de fiabilidad que presentan es elevado. Otras técnicas, como CCA y CDA, aunque presentan una buena fiabilidad tienen el inconveniente de que distribuyen mucho los puntos en la proyección dificultando la visualización de patrones en los datos.

Dado que se va a trabajar con vecindades bastante elevadas y se busca sobre todo garantizar que los puntos proyectados en el espacio de salida sean los más cercanos posibles en el espacio de entrada, se necesitan técnicas que posean una elevada fiabilidad. Cada técnica se usará para un propósito concreto:

- **SNE.** Se usará fundamentalmente para la generación de proyecciones con mapas continuos, ya que de las dos técnicas escogidas es la que mejor continuidad tiene. Esto permitirá que los mapas continuos generados con el GRNN no generen puntos con valores de tipo *NaN* y distribuyan los valores interpolados por todo el mapa.
- **t-SNE.** Se usará para generar los mapas de comparación de los edificios. Para generar estos mapas se requieren grandes vecindades con el fin de garantizar que la proyección de los datos tenga en cuenta todos los posibles puntos próximos. El único problema de usar esta técnica es que los mapas continuos presentan puntos de discontinuidad al agrupar demasiado los datos, por lo que las curvas de nivel de estos mapas suelen quedar muy próximas.

6.2.2. Proyección de los edificios

La comparación de perfiles va a permitir determinar el comportamiento del edificio en base a su consumo. En las figuras 6.9 y 6.10 se muestran los resultados de aplicar el SNE a los datos de Filosofía y Animalario respectivamente. Se puede observar cómo los días se ordenan en función de su consumo medio y dónde estarían las distintas zonas gracias a las curvas de nivel. En las subfiguras (b) y (c) de las figuras 6.9 y 6.10 se muestra la misma proyección pero el tamaño de los puntos corresponde al valor de la temperatura media de ese día y de la radiación solar, respectivamente.

La diferencia en la forma de la proyección da una idea del comportamiento del edificio. En Animalario, que es un edificio con poca dependencia de la hora y de un consumo más o menos constante, el mapa continuo adquiere una forma de abanico, mientras que en Filosofía el plano es más lineal. Esto indica que entre los puntos máximo y mínimo hay una mayor distancia en el caso de Filosofía que en el de Animalario. En las subfiguras (a), el radio de los puntos depende del índice de actividad del campus, lo que demuestra que el consumo de Filosofía es alto para aquellos días en los que el campus tiene actividad y bajo en el caso contrario; sin embargo, en Animalario se ve claramente como el consumo no depende en ningún momento de la actividad del campus. Para el caso de las variables ambientales, en el caso de Filosofía se puede apreciar cómo los días de menor radiación solar y de menor temperatura coinciden con los puntos de mayor consumo de potencia, mientras que para Animalario esto no es siempre así.

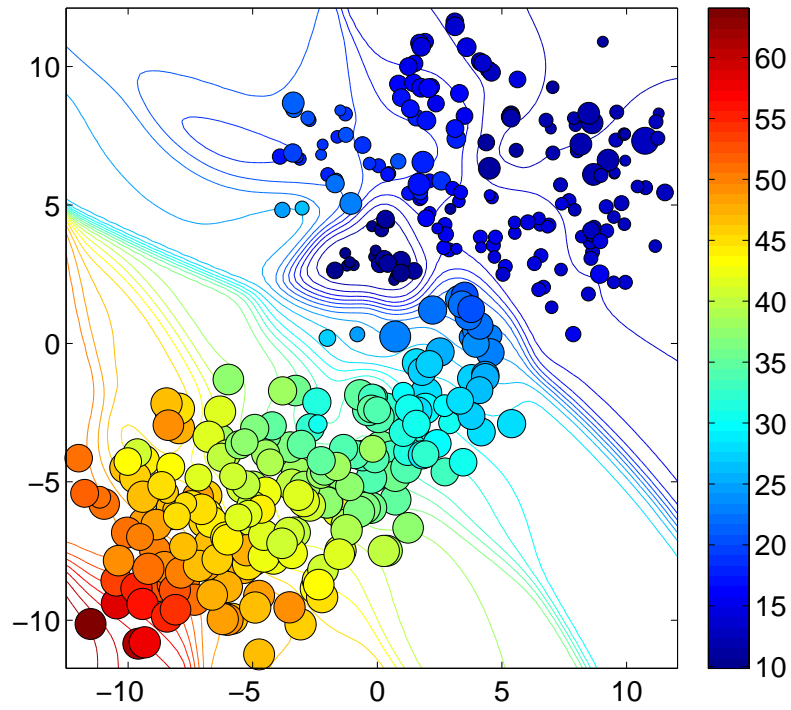
6.2.3. Comparación mediante técnicas de reducción de la dimensionalidad

La comparación de perfiles tiene como objetivo principal determinar qué edificios tienen un comportamiento similar de manera visual. La agrupación de los edificios será útil en experimentos posteriores, ya que simplificará los cálculos.

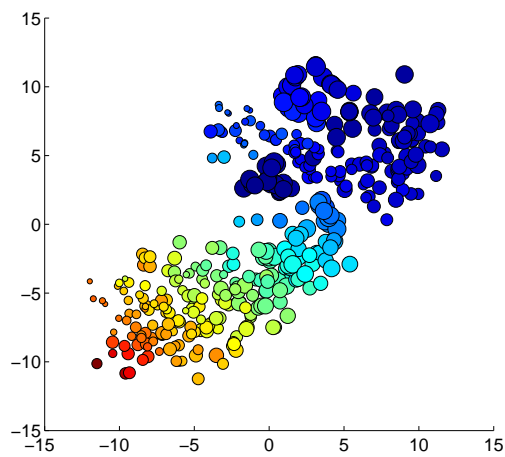
Dada la gran cantidad de edificios y las distintas posibilidades disponibles a la hora de comparar los consumos entre sí, se ha desarrollado una herramienta en *Processing* que permite interactuar con los datos y facilita su comparación. Se pueden realizar distintos tipos de visualizaciones para comparar edificios entre sí o bien para comparar distintos días dentro de un mismo edificio, o entre edificios. La figura 6.11 muestra una vista general de la herramienta, en la que aparecen varios edificios seleccionados simultáneamente.

En la herramienta el color de los puntos es dependiente del edificio al que pertenezcan, indicándose en la lista de edificios el color correspondiente de los puntos al seleccionarlo. El radio de los puntos es función de la variable que se haya seleccionado para visualizarlos, pudiéndose cambiar entre el índice de actividad, la temperatura, la radiación solar o la potencia media consumida en ese día. De esta forma se facilita la visualización de la relación entre las variables ambientales y los perfiles de consumo. Para el caso del índice de actividad se puede ver en la figura 6.12 que los puntos de menor radio se proyectan en la zona izquierda, mientras que los de elevada actividad se encuentran en la zona derecha.

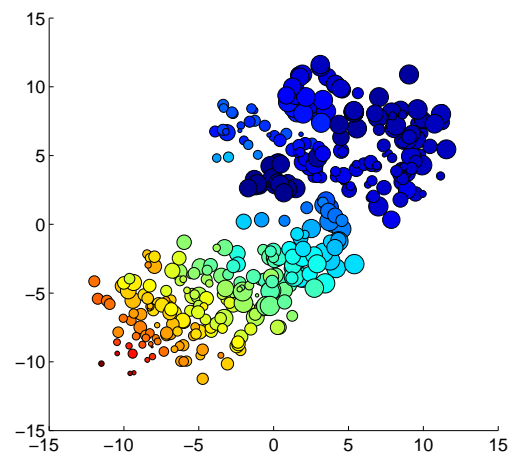
En cuanto a los diagramas de barras, se utilizan dos diagramas. El diagrama de la parte superior izquierda muestra el valor de las variables ambientales para el punto seleccionado.



(a) Proyección SNE en mapas continuos. Radio = Índice de actividad.

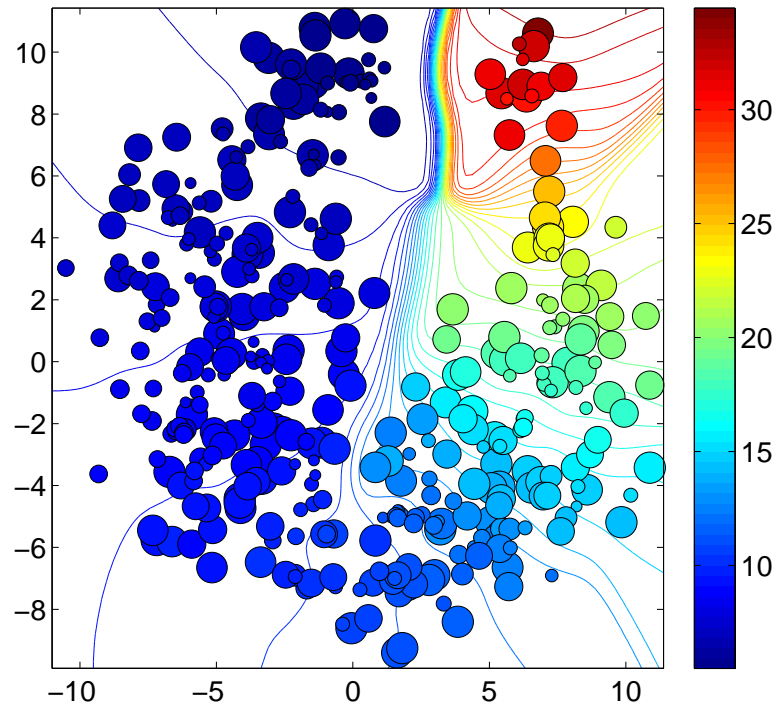


(b) Radio = Temperatura.

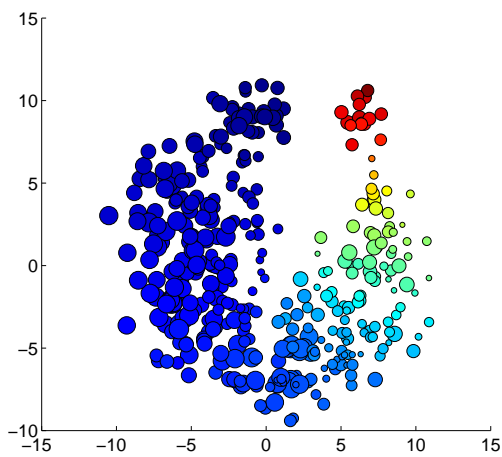


(c) Radio = Radiación solar.

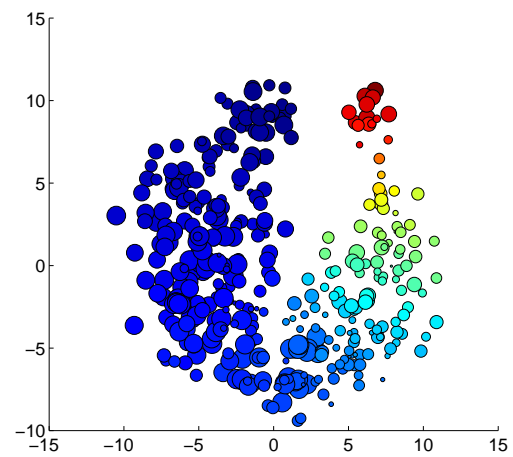
Figura 6.9: Proyección SNE de los perfiles de consumo para el edificio de Filosofía. El color indica la potencia media consumida y el radio una variable ambiental.



(a) Proyección SNE en mapas continuos. Radio = Índice de actividad.



(b) Radio = Temperatura.



(c) Radio = Radiación solar.

Figura 6.10: Proyección SNE de los perfiles de consumo para el edificio de Animalario. El color indica la potencia media consumida y el radio una variable ambiental.

El diagrama inferior es el perfil del consumo a lo largo del día para el punto seleccionado. En ambos gráficos el color indica la proximidad al valor máximo registrado para ese valor, así un color rojo indica valores cercanos al máximo y colores azules, valores cercanos al mínimo.

A la hora de comparar se han utilizado dos métodos distintos:

- **Proyección con SNE utilizando normalización común.** Se han proyectado todos los datos de todos los edificios, normalizados conjuntamente. Esta normalización permite comparar los perfiles diarios teniendo en cuenta la magnitud del consumo de cada edificio. Con esta proyección se quiere lograr obtener mapas de características de los datos a los cuales se les añade, mediante una interpolación obtenida con el GRNN, curvas de nivel que permiten una visualización continua de la distribución de los datos. Esto permite explorar la distribución de los puntos proyectados así como la visualización de zonas no contempladas en la reducción de la dimensión, como son las fronteras entre los días de elevado consumo con los días de baja actividad. Se utiliza el SNE que proporciona una mayor continuidad y genera interpolaciones con menor error. En la figura 6.11 se muestra el resultado de esta proyección para varios edificios. Se puede apreciar cómo algunos edificios se localizan cercanos, pero la mayoría quedan distribuidos por todo el mapa ya que la mayor parte de los edificios tienen distintos niveles de consumo.
- **Proyección con t-SNE utilizando normalización independiente.** En el caso de esta proyección, la normalización se ha hecho entre 0 y 1 para cada edificio. De esta forma, todos los edificios están en el mismo rango de valores. Esta proyección tiene como objetivo poder comparar la dinámica del perfil de cada edificio y no tanto el nivel de consumo de cada uno. De esta forma, comparando dinámicas, se podrán agrupar luego los edificios por dependencia horaria y no teniendo en cuenta el nivel de consumo, que no influye en la dependencia horaria. Dado que en este caso no interesa tanto generar una proyección continua, puesto que va a hacerse únicamente con el fin de poder obtener información acerca del comportamiento dinámico, se utiliza el t-SNE, que garantiza una mayor fiabilidad en las visualizaciones. En la figura 6.12 se muestra la proyección de los mismos edificios que en la figura anterior pero usando el t-SNE. Se puede ver cómo en esta proyección sí existen edificios que se agrupan, lo que indica que su perfil de consumo es similar.

La herramienta desarrollada en *Processing* permite hacer una transición entre las dos proyecciones con el fin de poder ver como se va desplazando cada uno de los edificios en función de su nivel de consumo o su dinámica. Dependiendo de lo que se quiera analizar se usará una proyección u otra. Esta transición permite ver el cambio que se produce en las posiciones en las que se proyectan los edificios y analizar cómo edificios que en principio tienen niveles de consumo diferentes en realidad luego presentan la misma dinámica y por tanto el mismo patron de consumo aunque el nivel de demanda sea diferente.

La herramienta de *Processing* también incorpora más funcionalidades como un calendario y una gráfica para la comparación de puntos (ver figura 6.13). El calendario permite

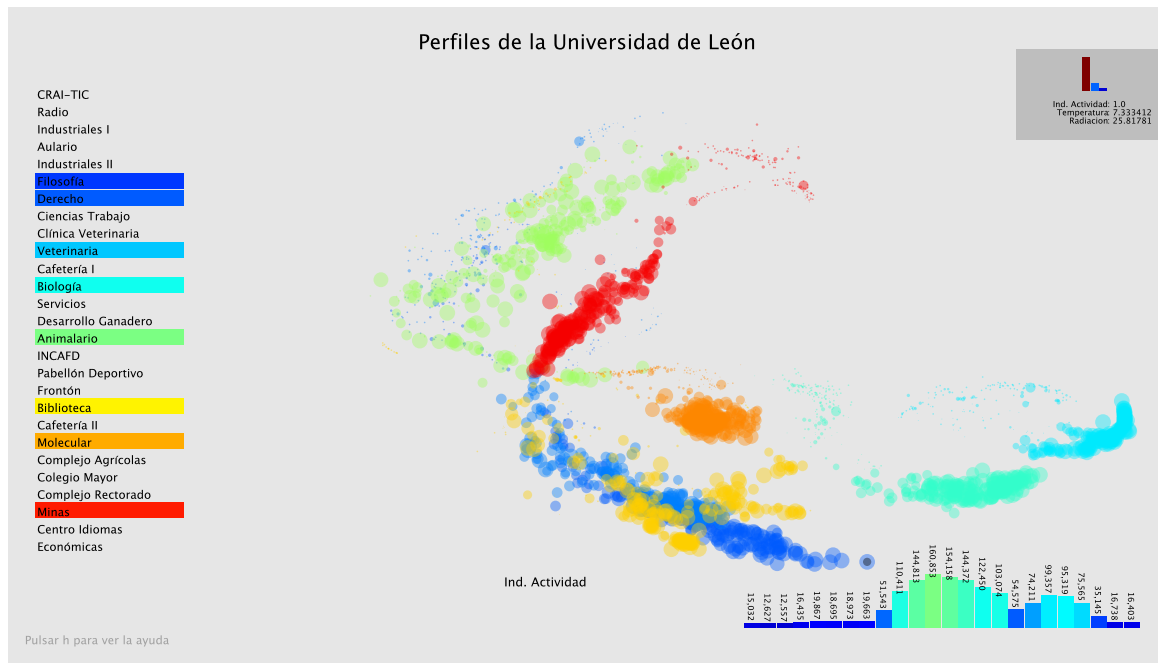


Figura 6.11: Comparación de perfiles de varios edificios usando SNE.

seleccionar los días cuyas curvas se quieren proyectar para poder realizar las comparaciones. Este calendario permite no solo la selección por días, sino que también permite seleccionar todo el año, todos los días de un mes concreto o todos los días que corresponden a un mismo día de la semana. Esto facilita la selección de días cuando se quieren comparar periodos de tiempo determinados, como por ejemplo ver la diferencia existente entre el consumo de los meses de verano y de invierno.

Además, la herramienta permite seleccionar varios puntos para poder comparar gráficamente los perfiles de cada uno de ellos. En el caso de la figura 6.13, se ha seleccionado un punto de cada uno de los edificios proyectados para poder ver como varía el perfil de un edificio a otro. Estos perfiles aparecen en una gráfica en la parte inferior de la aplicación, mostrando cada día seleccionado con un color diferente. De esta forma se pueden seleccionar distintos días y ver cómo cambia la dinámica del perfil en función del día.

A modo de ejemplo de comparación entre edificios, en la figura 6.12 se comparan varios edificios del campus, que presentan distintos tipos de consumo presentes en la Universidad y que se corresponden con un edificios de docencia, laboratorios y de servicios. En la proyección se puede ver claramente cómo los puntos correspondientes a los edificios de Filosofía, Derecho y Minas, que se corresponden a edificios de docencia, se proyectan todos agrupados en la misma zona y presentan una distribución prácticamente lineal. Otros edificios, como Biología, Veterinaria y Molecular, que tienen una actividad más orientada hacia la investigación, se proyectan en otra zona del mapa, adquiriendo una distribución más circular. Por último edificios como Biblioteca o Animalario, que son edificios de servicios o de apoyo, tienden a extenderse por toda la proyección. Esto se debe fundamentalmente a que el perfil de consumo de estos últimos es bastante plano, mientras que el de los otros

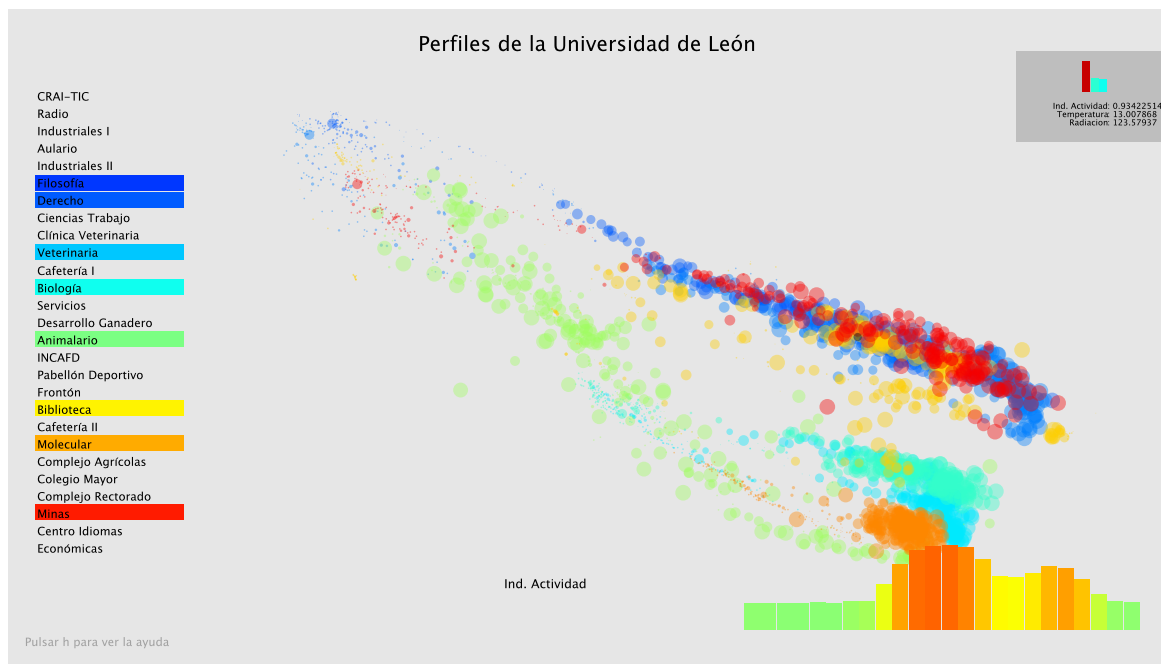
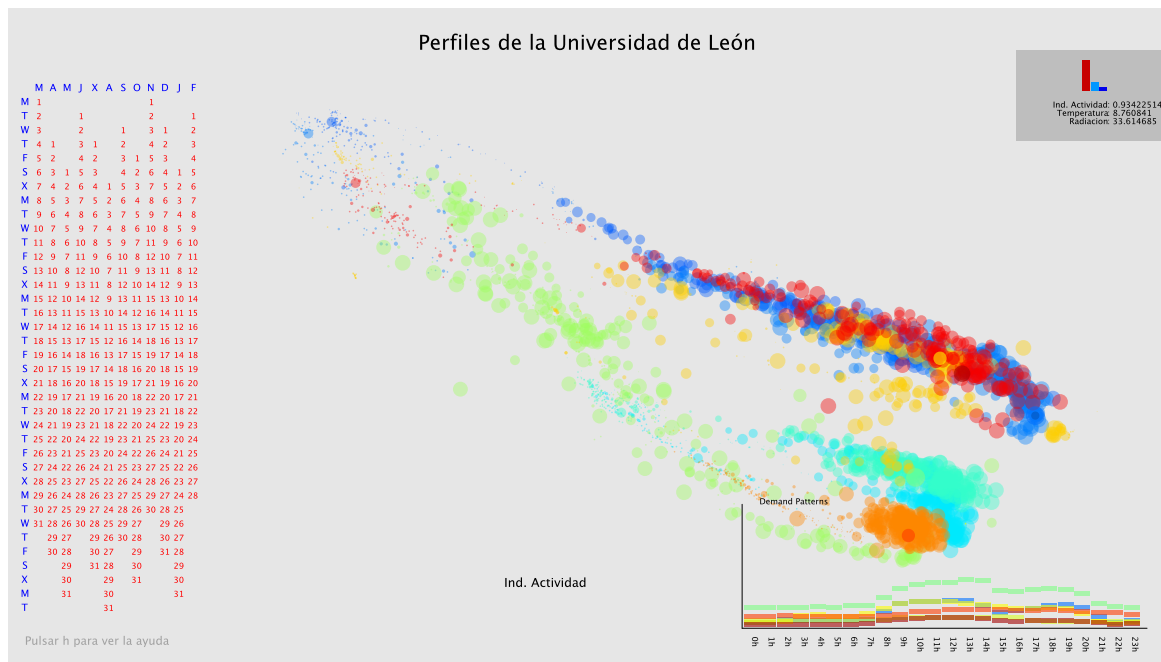


Figura 6.12: Comparación de perfiles de varios edificios usando t-SNE.

Figura 6.13: Calendario y gráfica de perfiles en *Processing*.

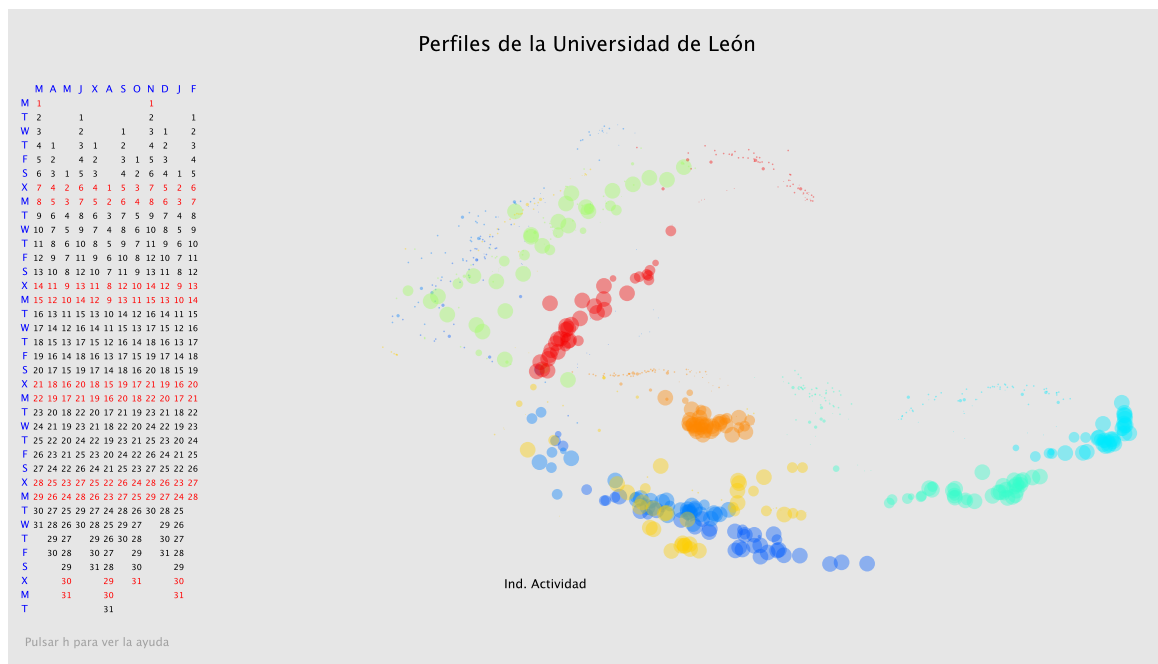


Figura 6.14: Comparación del consumo de edificios en función del día de la semana.

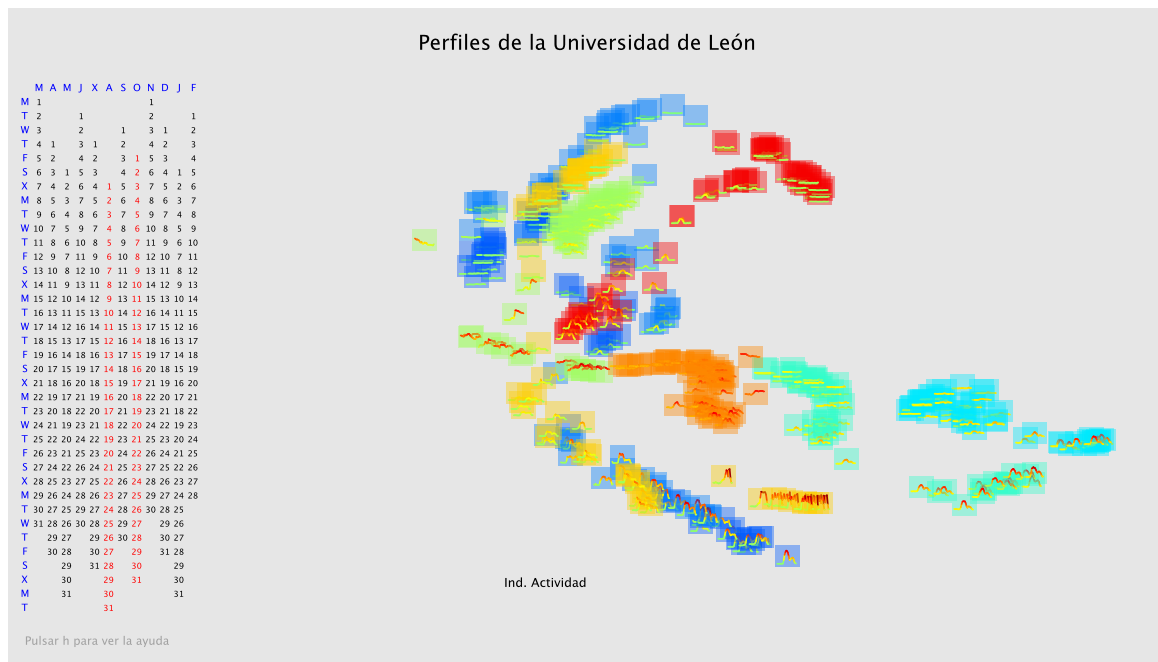


Figura 6.15: Comparación del consumo de edificios en función del mes.

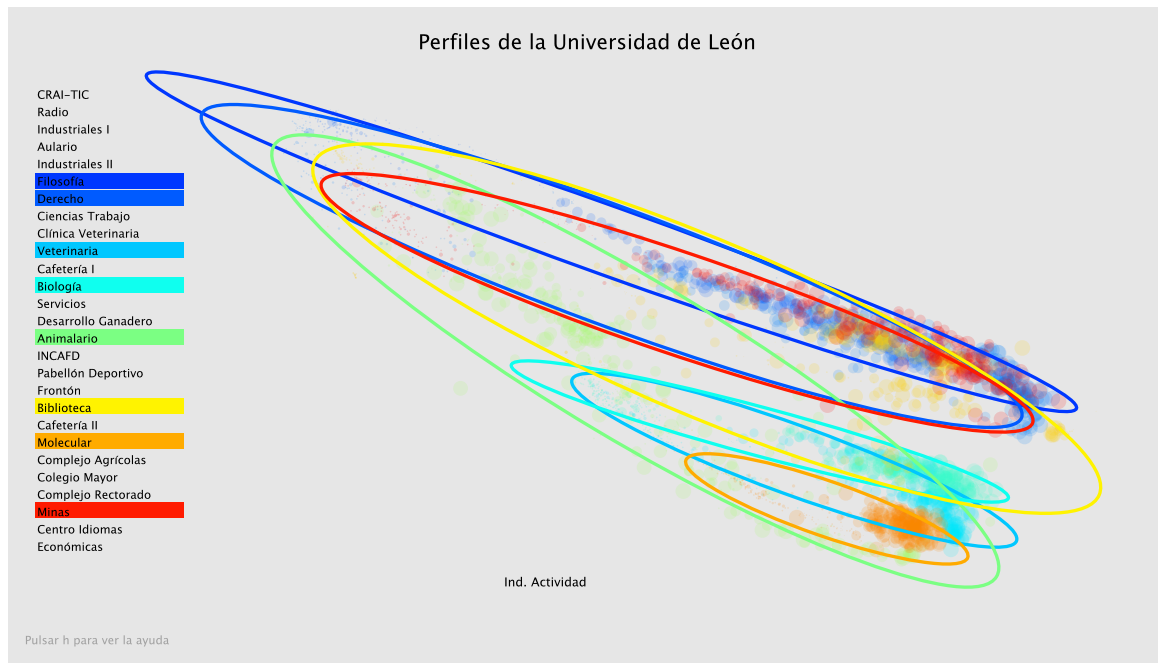


Figura 6.16: Comparación del consumo de los edificios en función de la elipse que les contiene.

dos edificios es mucho más dependiente de la hora. Los edificios de Biología y de Filosofía se separan porque el consumo base de ambos difiere entre sí.

Por último, se muestra la utilidad de comparar días entre sí. Las figuras 6.14 y 6.15 comparan los consumos entre los mismos días de la semana, en el caso de la primera figura, y entre meses, en el caso de la segunda. En la primera se puede ver cómo los días de menor actividad se agrupan todos en una zona, correspondiente a las zonas superiores del mapa, mientras que los lunes, que son días de mayor actividad, se encuentran en el extremo contrario, la parte inferior de la proyección. La figura 6.15 muestra, en vez de puntos, las curvas de consumo para cada día. Para esta proyección se ha seleccionado el mes de agosto, que es de baja actividad, y el mes de octubre, que es el caso contrario. Se puede ver como los días de baja actividad se corresponden con curvas planas y son días pertenecientes al mes de agosto, que se proyectan en la parte superior, mientras que los días de mayor actividad se corresponden con el otro mes seleccionado (salvo los fines de semana que son días de bajo consumo).

Clustering en base a las proyecciones de los perfiles dinámicos

Con la proyección t-SNE es posible hacer un *clustering* en base a la morfología de las proyecciones en el espacio de baja dimensión. Para ello, cada edificio se modeliza de acuerdo a la posición del centro de la elipse que contiene a todos los puntos proyectados y a la magnitud de sus ejes. De esta forma el vector que codifica el edificio es $E = (C_x, C_y, a, b)$ donde a y b son los semiejes de la elipse y las variables C_x y C_y son las coordenadas del centro. En la figura 6.16 se muestra un ejemplo de las elipses que contienen los puntos proyectados en *Processing*.

Grupo 1	Grupo 2	Grupo 3
CRAI-TIC	Industriales I	Veterinaria
Radio	Aulario	Biología
Clínica Veterinaria	Industriales II	Desarrollo Ganadero
Cafetería I	Filosofía	Molecular
Servicios	Derecho	Complejo Agrícolas
Animalario	Ciencias Trabajo	
Frontón	INCAFD	
Biblioteca	Pabellón Deportivo	
Cafetería II	Minas	
Colegio Mayor	Centro Idiomas	
	Económicas	
	Complejo Rectorado	

Tabla 6.4: *Clustering* de los edificios en base a la forma de la proyección de su dinámica.

Una vez obtenidos los vectores, estos se agrupan utilizando un *k-means*, dado que el número de puntos y *clusters* a generar es muy bajo y el *Neural Gas* para estos valores tiene una eficacia mucho menor a la hora de obtener mejores grupos. Utilizando el índice de *Davies-Bouldin* se obtiene que el número de *clusters* óptimo se encuentra entre 2 y 4 tipos de edificios, lo cual concuerda con la clasificación que se hizo en base a la utilización de los edificios en la sección 4.1. Por lo tanto, se agrupan los edificios en base a tres conjuntos que se recogen en la tabla 6.4.

Como se puede ver en dicha tabla, si la comparamos con la tabla 4.1 de los tipos de edificios en base a su utilización el grupo 1 se correspondería con edificios de apoyo, los del grupo 2 corresponden mayormente a facultades y los del grupo 3 a edificios de investigación. Si vemos, además, los datos de la tabla 6.1, podemos ver como los edificios del grupo 2 se corresponden con valores bajos de correlación horaria lo que corrobora que los edificios de este grupo tienen una alta dependencia de la potencia con la hora del día, los del grupo 3 con valores intermedios, con lo que la dependencia es inferior, y los del grupo 1 no tienen una dependencia tan clara con la hora del día.

6.3. Modelado de los edificios

Como se comentó en el capítulo 5, el SOM constituye un método de proyección especial, ya que incorpora una cuantización de vectores. Esta cuantización, dada por el número de neuronas con las que cuente el SOM, permite la reducción de los datos de entrada y trabajar con aquellos datos más representativos que hacen posible mostrar el comportamiento de las curvas de consumo para cada día y su relación con las variables ambientales.

Los métodos de proyección comentados anteriormente tienen el problema de que cuando el número de puntos es muy elevado, los algoritmos requieren de elevados tiempos de ejecución. Para el caso del t-SNE, el entrenamiento de la proyección de todos los edificios,

empleada en *Processing*, requirió más de 8 horas de ejecución en el servidor de explotación. Por lo tanto, en el caso de que se quieran comparar datos de edificios para un periodo de tiempo superior a un año, sería necesario generar modelos que reduzcan el número de datos, con el fin de que el tiempo de ejecución no sea desmesurado.

Los modelos generados por el SOM no solo permiten reducir la cantidad de datos con los que se trabaja, sino que pueden emplearse para hacer una comparación de edificios similar a la efectuada con la proyección t-SNE por medio de los mapas de diferencias. Además, estos modelos se usan para la predicción a largo plazo, como se verá en la sección 6.4.

6.3.1. Generación de modelos

Para la generación de modelos que permitan determinar el comportamiento de un edificio se utilizará el SOM. Estos modelos serán útiles en experimentos posteriores, como son la predicción de datos a largo plazo o la comparación de edificios por medio de mapas de diferencias.

Para generar estos modelos se utilizarán las variables ambientales seleccionadas por correlación, junto con el índice de actividad. Las variables utilizadas son las mismas que las utilizadas para las proyecciones, recogidas en la tabla 6.3. Al igual que se realiza con las proyecciones, se utilizan datos de un año (365 datos) y se genera un modelo por cada uno de los 27 edificios.

Con estas variables se componen los vectores de entrada para el modelo y se entrena un SOM para cada edificio con todos los datos disponibles. A la hora de entrenar el SOM, con el fin de lograr una mejor ordenación de los datos ambientales para poder visualizar la información de los perfiles de forma más sencilla, existen dos posibilidades:

- **Ponderar las variables ambientales.** Una vez normalizados los datos, se les da más peso a estas variables con el fin de que tengan más influencia en el proceso de aprendizaje del SOM.
- **Realizar un entrenamiento del SOM en dos etapas.** En la primera etapa se busca entrenar y ordenar los datos ambientales, por lo que solo se considerarían estas variables en el entrenamiento para luego, en una segunda fase, dejar fijas estas variables y entrenar únicamente las variables correspondientes a los perfiles.

En pruebas preliminares se utilizaron las dos versiones. Sin embargo, los resultados de comparar el modelo con los datos reales, fueron bastante similares, por lo que finalmente se optó por la primera opción, ya que computacionalmente es mucho menos costosa, al no ser necesario entrenar dos veces.

La solución ha sido entrenar un SOM de tamaño 20×15 , ya que la cantidad de datos es muy baja. Optar por mapas más pequeños conlleva que el modelo generado sea muy pobre, y si se usan más neuronas, se llega a la situación de tener más neuronas que datos. El número de épocas de entrenamiento ha sido de 500.

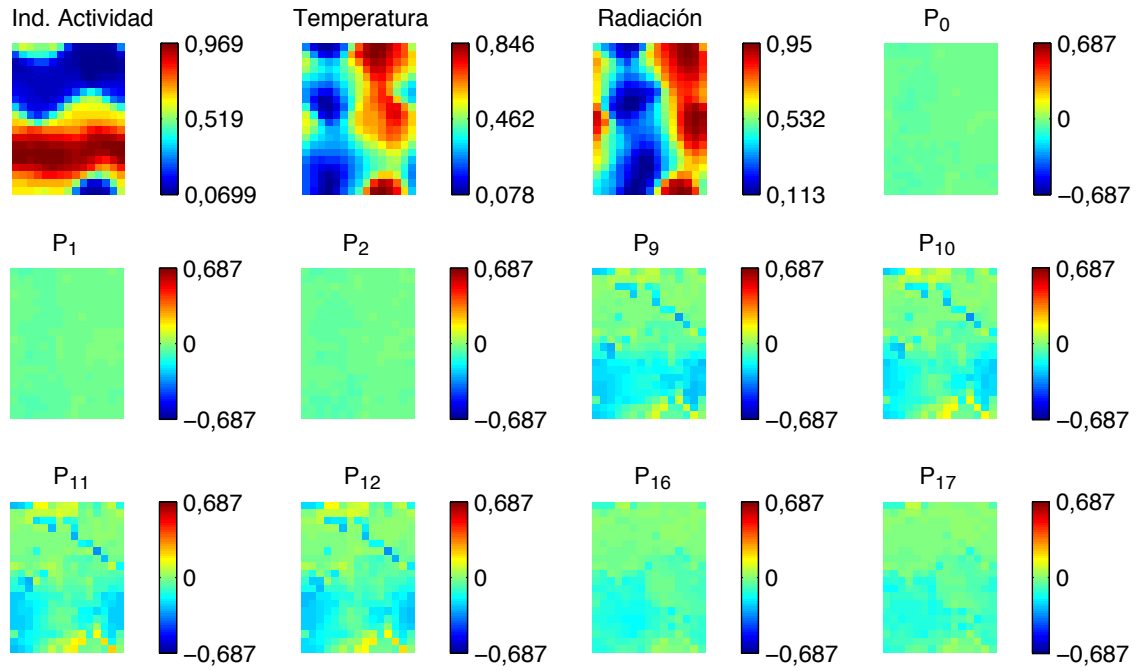


Figura 6.17: Mapas de diferencias para comparación de los modelos de los edificios de Filosofía y Derecho.

6.3.2. Comparación de modelos de perfiles

Con el fin de poder comparar los modelos obtenidos entre sí por medio del SOM, se plantea utilizar los mapas de diferencias explicados en la sección 5.3.2. Estos mapas indican las diferencias que se producen entre neuronas similares, ya que para cada neurona de un modelo se calcula la neurona más semejante en el otro modelo, siendo la diferencia entre ambas indicadora de la disimilitud producida entre los dos modelos.

Para aplicar estos mapas de diferencias, como ya se comentó, se trata de comparar aquellos días cuyas condiciones ambientales son similares con el fin de ver las variaciones de comportamiento entre los edificios en la misma situación, por lo que para obtener los modelos a comparar entre dos edificios, la neurona ganadora se calcula ponderando los datos ambientales de cada neurona en mayor medida frente a los consumos:

$$c_j = \arg \min_i \|\mathbf{w}_{Bj} - \mathbf{w}_{Ai}\|, \quad \text{donde } \mathbf{w} = [I_{ac}, T, R, P_0, P_1, \dots, P_{23}]. \quad (6.2)$$

Para comprobar este método se compara el edificio de Filosofía, que es el que presenta un perfil de consumo más correlacionado con la hora, con otro edificio con un perfil también dependiente de la hora, que en este caso es el edificio de Derecho, y con otro edificio de consumo no tan periódico y más constante a lo largo del día, que es el edificio de Animalario.

En la figura 6.17 se muestran los mapas de características para las variables ambientales y los mapas de diferencias para una serie de horas del día (no se incluyen todas con el fin de no saturar la visualización). Solo se emplean como horas del día una serie de horas representativas, que son tres horas de la noche (00, 01 y 02 h), tres horas del

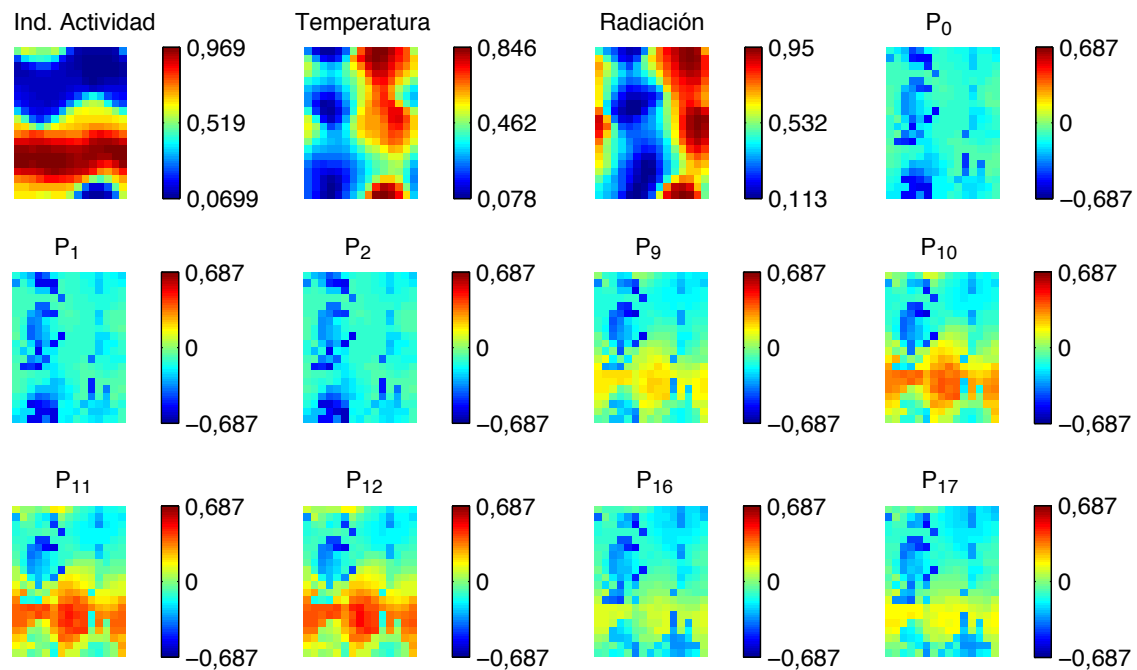


Figura 6.18: Mapas de diferencias para comparación de los modelos de los edificios de Filosofía y Animalario.

mediodía coincidentes con el máximo de consumo (09, 10, 11 y 12 h) y dos horas de la tarde coincidentes con el máximo global de consumo vespertino (16 y 17 h). Se puede ver claramente cómo aparecen pequeñas desviaciones a las horas punta de consumo los días de máxima actividad que se corresponden con las neuronas cercanas al rojo en el mapa de características del índice de actividad, siendo para el resto de horas insignificantes. Esto corrobora la hipótesis de partida de que ambos edificios, al corresponder su uso principalmente a clases, tienen patrones de consumo muy similares lo que se traduce en pocas desviaciones en los mapas de diferencias.

Por el contrario, en la figura 6.18, en la que se comparan los edificios de Filosofía y Animalario, se puede ver claramente cómo existen grandes diferencias para todas las horas del día, en especial en las horas de gran consumo. El color rojo de los mapas de diferencias marcan mayores valores en Filosofía, mientras que los azules marcan consumos más elevados en Animalario. Esto concuerda con los datos de partida que muestran que la evolución del consumo del edificio de Filosofía presenta grandes diferencias entre el máximo y el mínimo, mientras que la de Animalario es más "plana", siendo el consumo a lo largo del día más constante y presentando por ello menor dependencia horaria. Por lo tanto, estos mapas de diferencias nos muestran que los edificios de Filosofía y Animalario tienen usos completamente diferentes.

Para reducir el número de mapas necesarios para comparar los edificios, se usan mapas de módulos de diferencias que se explicaron en la sección 5.3.2. En el caso de que se usen estos mapas (ver figura 6.19) la visualización se reduce a un único mapa, lo que facilita la comparación entre los edificios. Así se puede ver más rápidamente como para el caso de

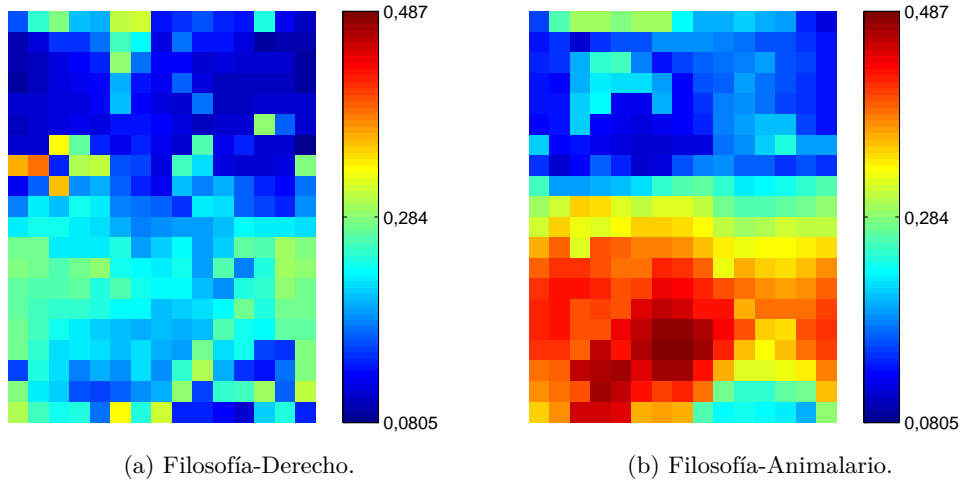


Figura 6.19: Mapas de módulos de diferencias.

Derecho, la diferencia global es mucho menor que en el caso de Animalario, sin necesidad de tener que comparar todos los mapas. Pero aun así, este mapa solo nos permite comparar dos edificios entre sí y no se puede hacer una comparación simultánea de todos los edificios.

6.4. Predicción a largo plazo

Los modelos de perfiles obtenidos para la comparación de edificios en el apartado anterior se van a utilizar para realizar la predicción a largo plazo. Considerando el método explicado en la metodología, los vectores de datos van a estar compuestos de dos partes:

$$\mathbf{x} = [I_{ac}, T, R, P_0, P_1, \dots, P_{23}] \quad \text{donde} \quad \begin{cases} \mathbf{x}^e &= \begin{bmatrix} I_{ac}, & T, & R \end{bmatrix} \\ \mathbf{x}^a &= \begin{bmatrix} P_0, & \dots, & P_{23} \end{bmatrix} \end{cases} \quad (6.3)$$

A la hora de calcular la neurona ganadora para determinar el perfil de consumo de ese día se utiliza únicamente la parte \mathbf{x}^e , la cual es conocida de antemano. Con este vector se calcula la neurona ganadora, siendo la parte \mathbf{w}^a de dicha neurona la que constituye los valores de potencia predichos a lo largo de ese día.

El índice de actividad (I_{ac}) se determina calculando el *cluster* ganador del *Neural Gas*, que fue entrenado según se indicó en la sección 6.1. Para ello, se codifica cada día del que se quiera hacer la estimación mediante un vector con las variables de la tabla 6.2, cuyos valores se asignarán acordes al calendario escolar de la Universidad. En cuanto a las variables meteorológicas, se podían obtener bien utilizando modelos disponibles en Internet o en base a los propios modelos creados. Las variables ambientales se calcularán de la siguiente forma:

- **Temperatura.** Se ha obtenido por medio de páginas de Internet de predicción meteorológica.

- **Radiación solar.** Dado que no se ha encontrado ningún modelo que proporcione una estimación de la radiación solar que esté accesible fácilmente a través de Internet, los datos para esta variable se obtendrán a partir de valores de días similares.

La predicción ha sido probada para los edificios de Filosofía, que es de carácter periódico, y de Animalario, que no presenta una regularidad en sus consumos. De esta forma se puede comparar un edificio de comportamiento periódico con uno que no lo es. El error de estos modelos se ha calculado con los índices recogidos en la sección 5.4 y se ha utilizado la técnica de validación cruzada *k-fold*, que consiste en subdividir el grupo de datos en *k* subconjuntos de los cuales uno se deja fuera del entrenamiento y es utilizado para validar, repitiéndose el proceso para todos los subconjuntos. El número de subconjuntos en este caso ha sido de 10, que es la partición más utilizada en la literatura. En la salida se utiliza una RBF para suavizar la predicción, la cual se ha entrenado con un radio de 0,15, ya que en la bibliografía se recomienda un valor cercano a 0,1 y en las pruebas preliminares es el valor que generaba menor error.

En las figuras 6.20 y 6.21 se muestran las predicciones que se realizan en base a estos modelos para dos edificios. La línea azul representa el consumo real que se ha producido en ese edificio, la línea roja continua es la predicción de la curva obtenida para las condiciones meteorológicas que se corresponden con la estimación más probable obtenida para ese día, y la línea roja y verde discontinuas muestran los márgenes de predicción tras modificar las variables ambientales a otras posibles situaciones. En cuanto a las desviaciones, se han calculado mediante el estudio de valores pasados de las variables ambientales para esos días concretos. Calculando lo que se han desviado las variables ambientales con respecto a la media para ese tipo de día, se ha establecido cuál sería la situación ambiental más desfavorable (rojo) y más favorable (verde). Mediante este sistema de triple predicción se obtiene una banda de consumo en la cual se podría mover el consumo para esos días.

Como se puede ver en las figuras, la precisión de la predicción de este método es bastante baja. Además, existe un problema derivado del uso del SOM dado que, al hacer una media de los datos locales, este tiende a suavizar las puntas de consumo que se producen. Sin embargo, el objetivo de este método no es tanto obtener una predicción precisa, ya que para eso existen otros métodos de predicción a corto plazo, sino dar una estimación de consumo que permita conocer de antemano cuáles pueden ser los días más desfavorables, y por lo tanto poder hacer una planificación en aquellos días en los que exista un riesgo de superar el valor de pico de potencia contratada.

En el caso de la predicción de consumo en Filosofía, dado que el comportamiento del edificio es muy dependiente de la hora, se puede ver como el umbral de predicción contiene a dicha curva, por lo que sería viable el uso de esta técnica en este edificio. Sin embargo, en el caso del edificio de Animalario (ver Fig. 6.21) este método se torna completamente ineficaz, puesto que este edificio carece completamente de dependencia horaria. Por tanto, realizar una predicción a largo plazo para este edificio con este tipo de método sería imposible. Además se puede ver como para dicho edificio en algunos casos (entre las horas 100-120 aproximadamente) las predicciones de consumo para las condiciones más desfavorables son inferiores a la predicción en la que se conocen las variables ambientales.

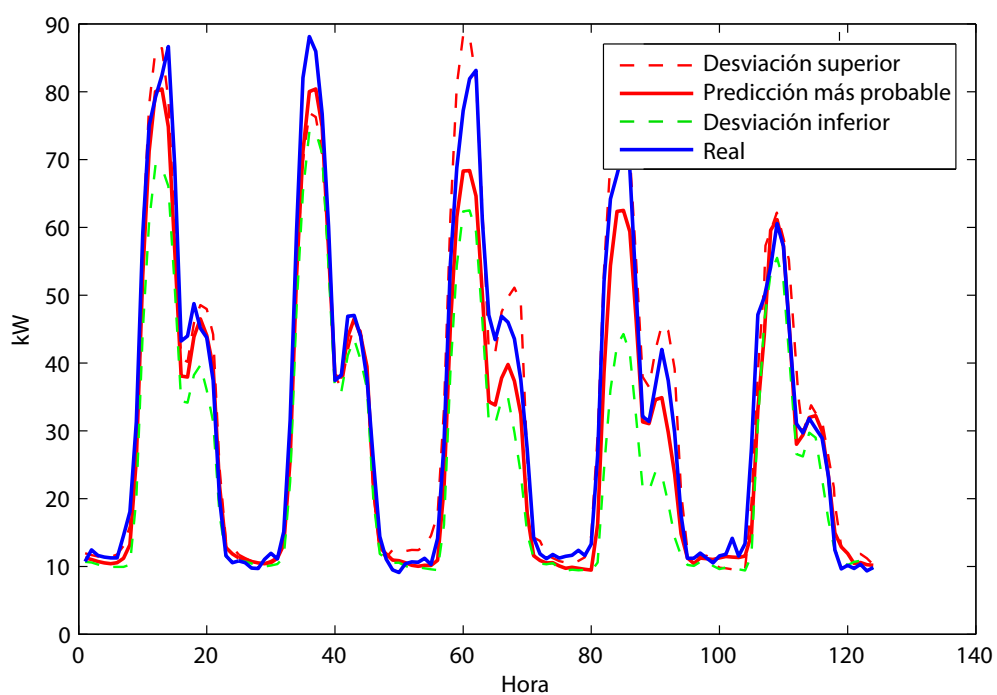


Figura 6.20: Predicción de consumos a largo plazo para el edificio de Filosofía.

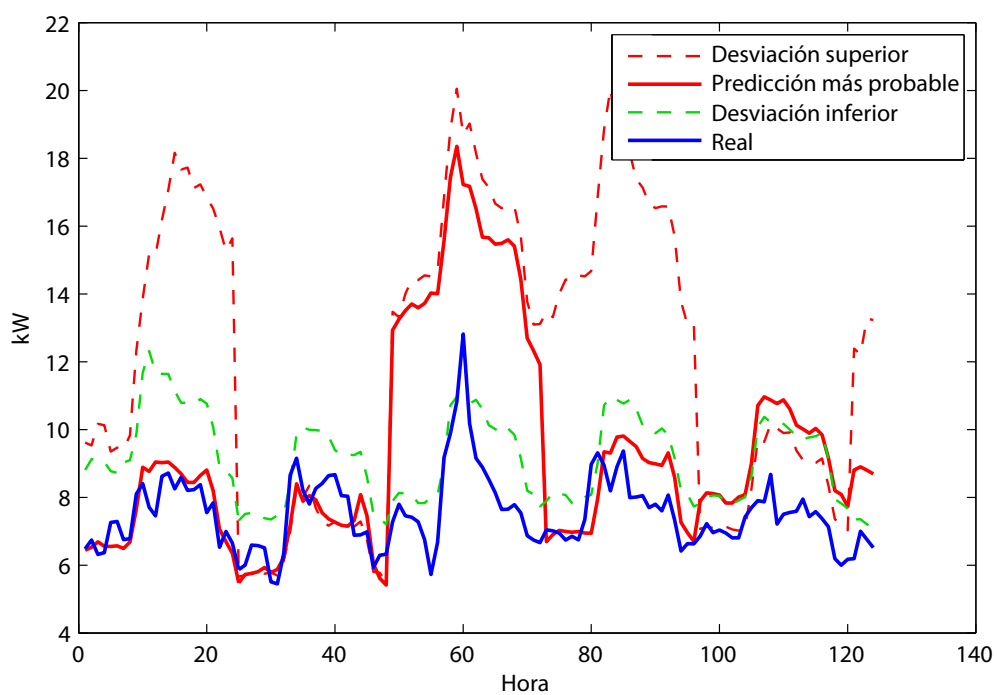


Figura 6.21: Predicción de consumos a largo plazo para el edificio de Animalario.

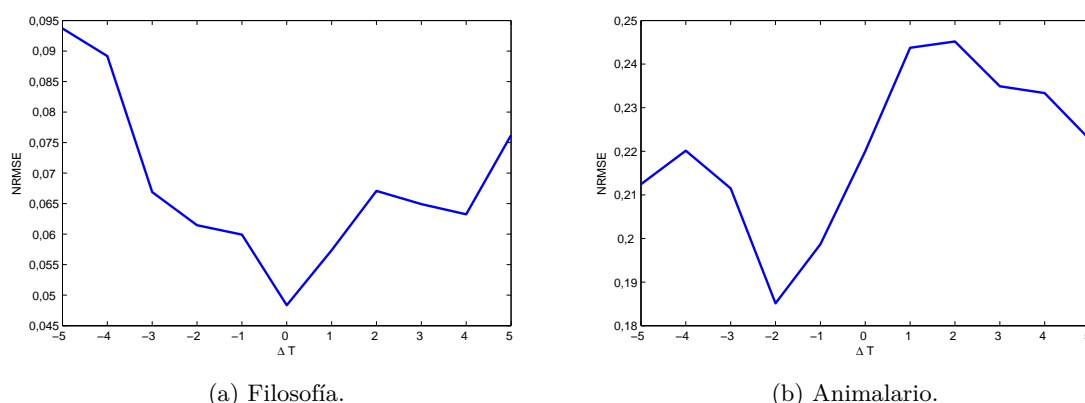


Figura 6.22: Comparación del NRMSE en función de la variación de la temperatura con respecto a la real.

En la figura 6.22 se muestra la variación la precisión de la predicción para los dos edificios en función de la desviación de la temperatura predicha con respecto a la temperatura media real para ese día. En el caso de Filosofía, el error aumenta conforme la temperatura predicha se desvía de la temperatura real; sin embargo, para el caso de Animalario esto no es así, generándose errores menores para desviaciones altas de la temperatura.

La tabla 6.5 refleja los índices de predicción MAPE, NRMSE y Theil U para todos los edificios, usando los modelos generados con el SOM. Se usan estos tres indicadores porque así también se puede ver cómo el uso de uno u otro puede en algunos casos conducir a conclusiones equivocadas:

- **MAPE.** Es el indicador más utilizado en la bibliografía para medir la precisión de una predicción de consumos eléctricos. Sin embargo, en todos los casos se usa para grandes consumos de energías ($P \simeq 1\text{GW}$), siendo completamente inútil para valores pequeños. Se puede ver cómo, para el caso de Frontón, el error del MAPE es superior al 1600 %. Esto se debe a que el consumo de Frontón en algunos momentos es cercano a cero, lo que implica que al calcular el MAPE hay términos divididos por cero. A la hora de comparar edificios entre sí tampoco es fiable, ya que hay casos de edificios que teniendo un MAPE menor, tienen un mayor NRMSE debido a que el nivel de consumo es muy superior.
- **NRMSE.** En el caso de edificios de consumo medio es el mejor indicador, ya que solventa el problema de los términos que dividen por cero que tiene el MAPE. Además, al estar normalizado entre cero y uno, permite comparar la precisión de la predicción entre los edificios. Así, tenemos que el edificio para el que funciona mejor la predicción es Filosofía, siendo el peor es Animalario. Si comparamos esta tabla con las correlaciones horarias de la tabla 6.1 vemos cómo, para el caso de edificios con alta correlación horaria (valor pequeño en la correlación mínima horaria), este método obtiene mejor resultado que para los edificios que no presentan dicha correlación.

ID	Edificio	MAPE	NRMSE	ThU
E1	CRAI-TIC	26,845	0,193	2,921
E2	Radio	39,933	0,107	1,515
E3	Industriales I	23,316	0,083	0,601
E4	Aulario	33,327	0,079	0,741
E5	Industriales II	18,337	0,092	0,864
E6	Filosofía	23,080	0,051	0,589
E7	Derecho	31,885	0,084	0,630
E9	Ciencias Trabajo	34,727	0,086	0,711
E10	Clínica Veterinaria	37,738	0,101	1,263
E11	Veterinaria	17,186	0,087	1,007
E12	Cafetería I	62,890	0,135	0,944
E13	Biología	11,312	0,059	0,639
E14	Servicios	22,743	0,072	0,640
E15	Desarrollo Ganadero	23,967	0,115	0,888
E16	Animalario	60,224	0,157	2,867
E17	INCAFD	34,623	0,077	0,755
E18	Pabellón Deportivo	66,476	0,098	0,944
E19	Frontón	1655,330	0,164	0,938
E20	Biblioteca Central	67,691	0,142	1,236
E22	Cafetería II	211,793	0,083	1,435
E23	Molecular	8,641	0,063	0,743
E24	Complejo Agrícolas	18,139	0,083	0,885
E25	Colegio Mayor	32,058	0,086	1,067
E26	Complejo Rectorado	35,447	0,118	1,132
E28	Minas	28,229	0,070	0,633
E29	Centro Idiomas	28,336	0,078	0,628
E31	Económicas	22,028	0,082	0,591

Tabla 6.5: Índices de predicción para todos los edificios de la universidad usando los modelos generados con el SOM.

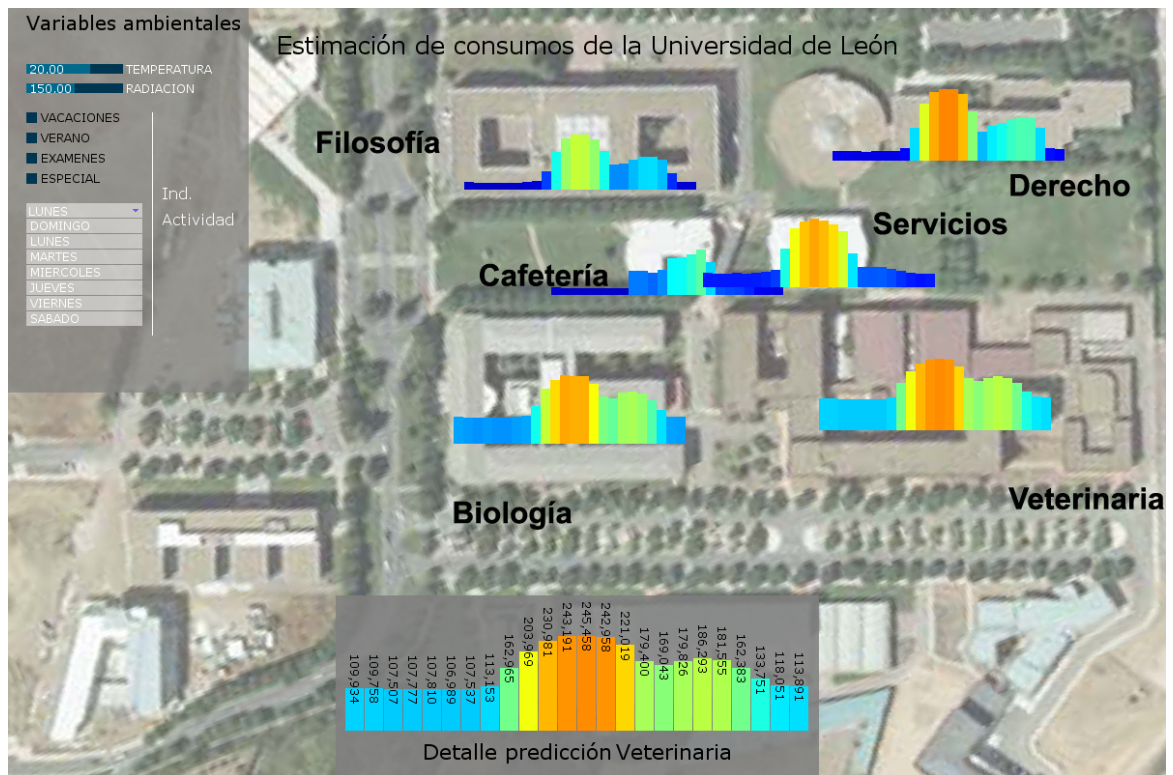


Figura 6.23: Herramienta de *Processing* para la visualización de la predicción de consumos.

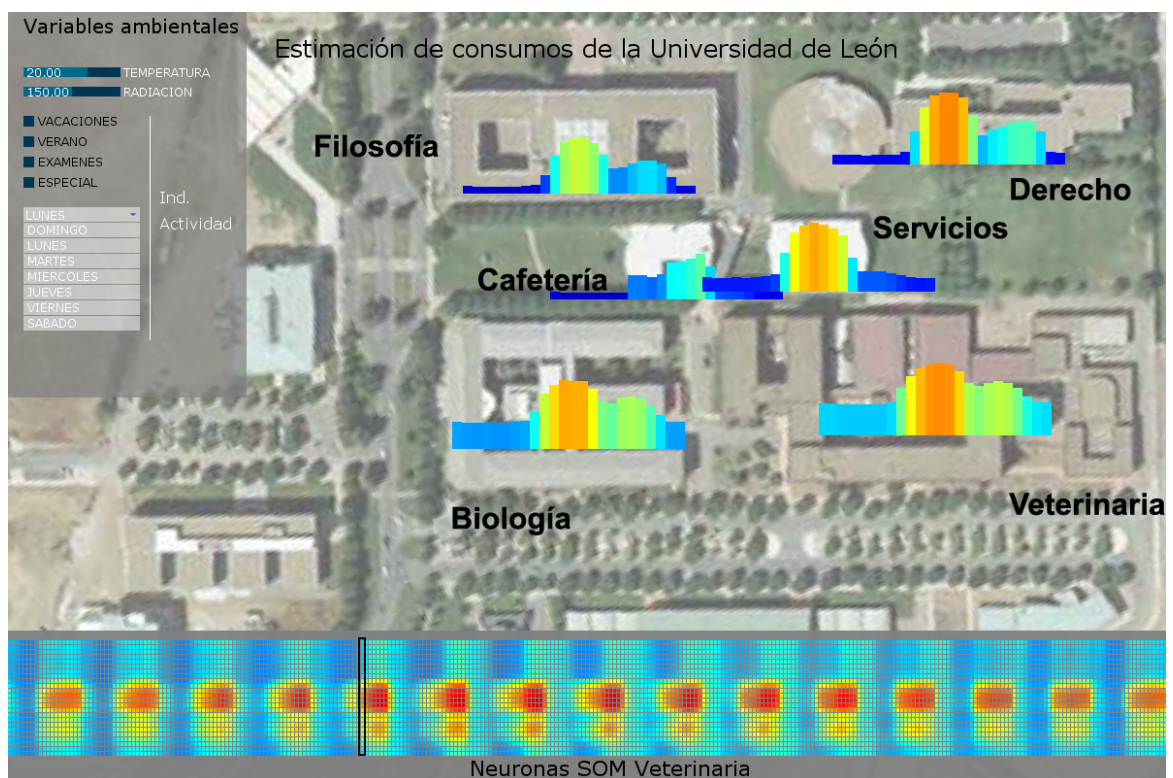


Figura 6.24: Herramienta de *Processing* para la visualización de los modelos del SOM.

- **Theil U.** Este índice pretende mostrar si la predicción es mejor que utilizar un método simple ($ThU < 1$). En este caso, el método simple sería utilizar como predicción para un día el valor del mismo instante de tiempo del día anterior, ya que en la predicción se obtiene el perfil diario. En la tabla 6.5 se puede ver como hay edificios que no son periódicos y en los que el método de predicción resulta mucho peor que considerar la predicción como el mismo consumo que el día anterior.

Los modelos para predicción a largo plazo permiten, por medio de una herramienta implementada en *Processing* (ver figura 6.23) hacer una estimación interactiva para todos los edificios de la Universidad de León. Por medio de una interfaz de operación se pueden seleccionar manualmente los parámetros ambientales que se esperan para un día en concreto. En función de esos parámetros, se predice una curva de consumo para esas condiciones ambientales. La posibilidad de cambiar los valores ambientales permite volver a calcular las curvas estimadas, con el fin de poder ver los cambios que se producen en el consumo en función de dichos parámetros. La herramienta incorpora la posibilidad de cambiar los parámetros ambientales para cada hora concreta del día, para poder estimar el consumo en el caso de días en los que no se conozcan con exactitud las condiciones y se prevean cambios a lo largo del día. Un ejemplo de esta situación es el caso de edificios de tipo docente, en los cuales la celebración de juntas que implican la suspensión de las clases o el caso en el que hay suspensión de clases por inicio de las fiestas propias de la facultad.

La herramienta incorpora también la visualización de los modelos generados con el SOM que son los que permiten realizar la predicción a largo plazo (ver figura 6.24). Esta visualización muestra la matriz de pesos de las neuronas, siendo las columnas cada neurona, con el fin de poder ver la distribución de los modelos. Además esta visualización permite observar obtener los mapas de diferencias entre los modelos para comparar los comportamientos de los edificios entre sí.

6.5. Predicción a corto plazo

Para la predicción a corto plazo se va a determinar cuál de los métodos resulta más eficaz para cada uno de los tipos de edificio que se obtuvieron en el *clustering* de la proyección. Dado que este agrupa los edificios en función de su comportamiento dinámico, el método que resulte eficaz para el edificio seleccionado es el que resultará útil para el resto de edificios de dicho *cluster*.

Para los edificios del *cluster* 1 se ha seleccionado como representativo el edificio de Animalario, para el *cluster* 2 el edificio de Filosofía y para el *cluster* 3, Veterinaria. Primeramente se va a seleccionar para cada tipo de edificio el regresor óptimo para la predicción directa, ya que el método MIMO no requiere de ninguna selección previa. En todos los casos, el método para la selección de los parámetros será una validación cruzada, utilizándose como método de validación el *10-fold*.

Para la generación de los modelos de predicción a corto plazo se utilizan las cuatro técnicas descritas en el capítulo anterior. El uso de estas técnicas combinan la utilización de los regresores de potencia con una serie de variables ambientales y de evolución de consumo

Etiqueta	Descripción
H_x	Hora del día a partir de la cual se predice. Coordenada x
H_y	Hora del día a partir de la cual se predice. Coordenada y
I_{ac}	Índice de actividad estimado para las siguientes 24 horas
\bar{T}_0	Temperatura media predicha para las siguientes 24 horas
\bar{R}_0	Radiación solar media predicha para las siguientes 24 horas

Tabla 6.6: Variables ambientales utilizadas para la predicción.

que permiten acotar el tipo de día a predecir. Las variables seleccionadas para todos los métodos son las mismas y se recogen en la tabla 6.6. Estas variables son las mismas variables ambientales que se seleccionaron para la proyección y se añade la información de la hora, con el fin de marcar el punto a partir del cual se va a realizar la predicción para las siguientes 24 horas.

En función del método utilizado, este tendrá un modo de entrenamiento. En todos los casos se busca predecir, a partir de un instante de tiempo, las siguientes 24 horas de consumo. Los regresores utilizados siempre se crean desplazando la ventana una hora, de tal forma que independientemente de la hora del día que sea, se puede conocer el consumo todo un día en adelante.

El número de datos que se utilizan son los correspondiente a 396 días y, por lo tanto, la entrada de cada método de predicción está compuesta por 9504 vectores. De estos datos se extrae información referente a cuatro semanas (768 muestras) de diferentes meses, que nunca se utilizará para entrenar y que es utilizada para poder comparar los resultados de cada método aplicado a cada edificio y que constituyen los datos de test.

6.5.1. Predicción MIMO

La predicción MIMO permite predecir las siguientes 24 horas de consumo en un único paso, obteniéndose de esta forma el perfil completo. A continuación se recogen los parámetros de entrenamiento de los métodos de salida múltiple utilizados.

Predicción MIMO con SOM

Para la predicción por medio del SOM, los datos de entrada de entrenamiento constan de tres partes recogidas en la siguiente ecuación:

$$\mathbf{x} = \left[\left(H_x, H_y, I_{ac}, \bar{T}_0, \bar{R}_0 \right), \dots, \left(P(t-24D+1), P(t-24D+2) \right), \dots, P(t-1), P(t) \right), \dots, \left(\hat{P}(t+1), \hat{P}(t+2), \dots, \hat{P}(t+24) \right) \right] \quad (6.4)$$

1. Parte ambiental, compuesta por las variables de la tabla 6.6.
2. Parte regresor, compuesta por los valores de potencias de los D días anteriores para predecir.

3. Parte predicción, que son los valores de potencia de las 24 horas a predecir.

Para entrenar el SOM, el método seleccionado es el VQTAM modificado, en el que se utiliza la siguiente máscara durante el entrenamiento:

$$\Omega = [(0, 0, 0, 0, 0), (1, 1, \dots, 1, 1), (0, 0, \dots, 0)]. \quad (6.5)$$

A la hora de calcular la neurona ganadora para la predicción, sí se tendrá en cuenta la parte ambiental.

Para determinar el tamaño del regresor, se utiliza la validación cruzada, obteniéndose valores de D en torno a 5 para edificios como Filosofía y Veterinaria, y de 2 para edificios como Animalario. Esto lleva a la conclusión de que cuanto menos dependiente de la hora del día sea el consumo de un edificio, menor será la longitud del regresor. En cuanto a las dimensiones del SOM, se ha optado por heurística por un tamaño que sea el 20 % de los datos de entrenamiento, lo que genera un modelo de unas 2000 neuronas, y dado que se obtienen mejores resultados con SOM rectangular, se utiliza un tamaño de 50×40 .

Predicción MIMO con CCA

En este caso, al igual que ocurría con los modelos obtenidos con el SNE y el t-SNE, se va a realizar una reducción de la dimensionalidad utilizando únicamente el perfil de consumo de 24 horas. El perfil a utilizar será el de las 24 horas que se quieren predecir y una vez obtenido el modelo de baja dimensión, se le asocia el conjunto correspondiente de variables ambientales.

Una vez que se tienen todos los datos correspondientes a tramos de 24 horas, se calcula la dimensión intrínseca de los datos para conocer cuál es la proyección más adecuada. Para el cálculo de esta dimensión, existen varios métodos, pero todos ellos proporcionan valores para el conjunto de datos disponible de entre 2 y 4, por lo que para simplificar los cálculos y seguir manteniendo la posibilidad de visualización de los datos, se ha optado por utilizar una dimensión de reducción de 2.

Una vez que se tiene la proyección, se calcula la trayectoria seguida por los datos en el espacio de baja dimensión y se predice cada dimensión por separado utilizando un modelo AR. Por medio de técnicas de validación cruzada se determina que el número de parámetros para la parte autorregresiva es de 10.

6.5.2. Predicción Directa

En el caso de predicción directa es necesaria una selección previa del regresor, ya que para cada hora se va a utilizar uno diferente con el fin de reducir el error. Una vez que se ha seleccionado el regresor, se entrena cada uno de los métodos de predicción explicados en la sección 5.5.2.

Selección del regresor

Para la selección del regresor se va a tener en cuenta la dependencia horaria del consumo del edificio; por ello, en función de esta dependencia, la búsqueda del regresor será diferente. Esta dependencia viene dada por la correlación calculada en la tabla 6.1. El regresor puede contener la siguiente información:

- **Parte secuencial.** Es aquella que está compuesta únicamente por información del día previo. Se restringe la búsqueda de la parte secuencial del regresor a las 24 horas previas, por tanto $M_1 = 24$ (ecuación 6.6). Todos los edificios tendrán información de la parte secuencial. La búsqueda se realizará por el método de filtrado por k -NN explicado en la sección 5.5.2. Será necesario determinar los k vecinos necesarios, lo que se hará por medio de una validación cruzada usando un *10-fold*. Los métodos de búsqueda serán tanto hacia delante como hacia atrás. Aquel regresor que genere un menor error de los ganadores de cada método será el seleccionado para predecir:

$$\hat{P}(t+s) = f[P(t), P(t-1), \dots, P(t-(M_1-1))]. \quad (6.6)$$

- **Parte periódica.** Se utiliza información de la misma hora de días anteriores. Se restringe la búsqueda de días anteriores a seis, ya que en las pruebas preliminares tamaños superiores no dieron buenos resultados, por tanto $M_2 \leq 6$ (ecuación 6.7). Este regresor es útil para edificios con una elevada dependencia horaria. Para determinar el número de días a tener en cuenta se utiliza el método *wrapper*, empleando una validación cruzada:

$$\hat{P}(t+s) = f[P(t+s-24), P(t+s-48), \dots, P(t+s-24M_2)]. \quad (6.7)$$

Dado que no hay ningún edificio que sea completamente dependiente de la hora del día, los edificios con alta y media dependencia horaria usarán **regresores mixtos** que quedarán de la siguiente manera:

$$\hat{P}(t+s) = f[P(t), P(t-1), \dots, P(t-(M_1-1)), P(t+s-48), \dots, P(t+s-24M_2)] \quad (6.8)$$

donde los valores de M_1 y M_2 son los que se indicaron anteriormente. Para aquellos edificios que contengan parte periódica, primero se realiza la búsqueda de dicha parte, puesto que es la que más reduce el error de predicción. De esta forma, por validación cruzada, se va añadiendo información de las horas similares hasta que el error de predicción empieza a aumentar. Una vez que se ha determinado el tamaño de la parte periódica, se añade la información de la parte secuencial partiendo del regresor periódico. El regresor que por k -NN genere un menor error será el seleccionado. Para el caso de edificios con baja dependencia horaria se utilizarán **regresores secuenciales**.

Los edificios del *cluster* 1 de la tabla 6.4 son edificios que tienen muy poca dependencia horaria, por lo que para generar el regresor únicamente se va a hacer una búsqueda en la parte secuencial. En la figura 6.25 se muestra el resultado de la búsqueda del regresor. Los cuadrados negros indican las muestras que son consideradas en cada regresor.

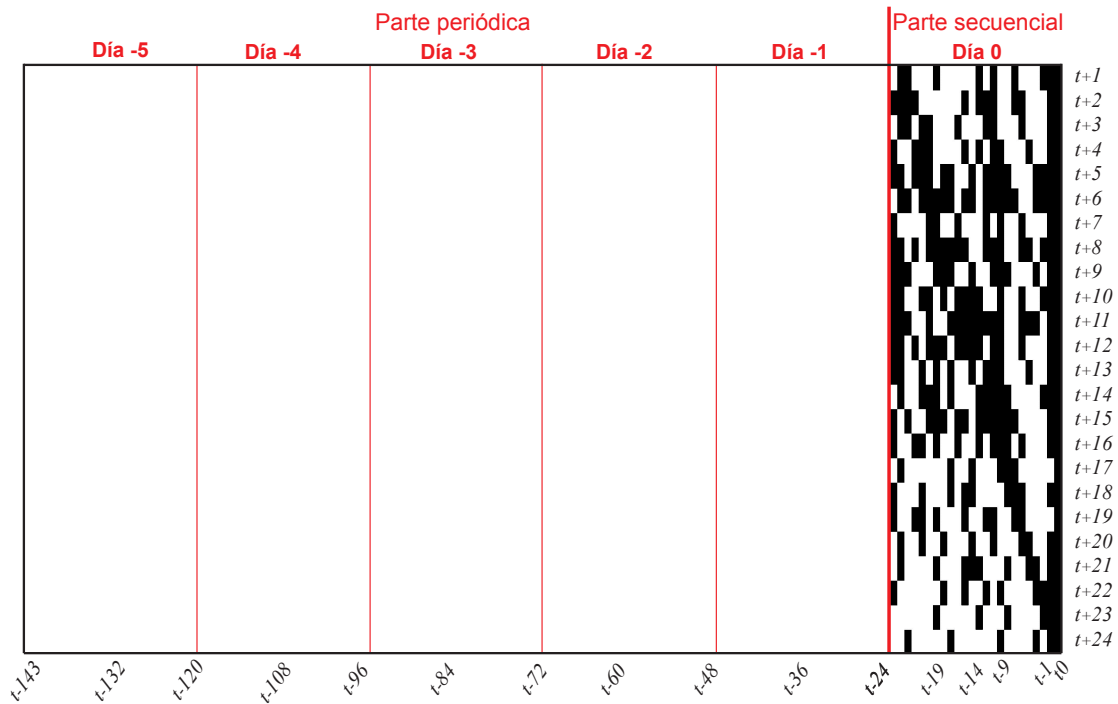


Figura 6.25: Selección del regresor para predicción directa correspondiente al edificio de Animalario.

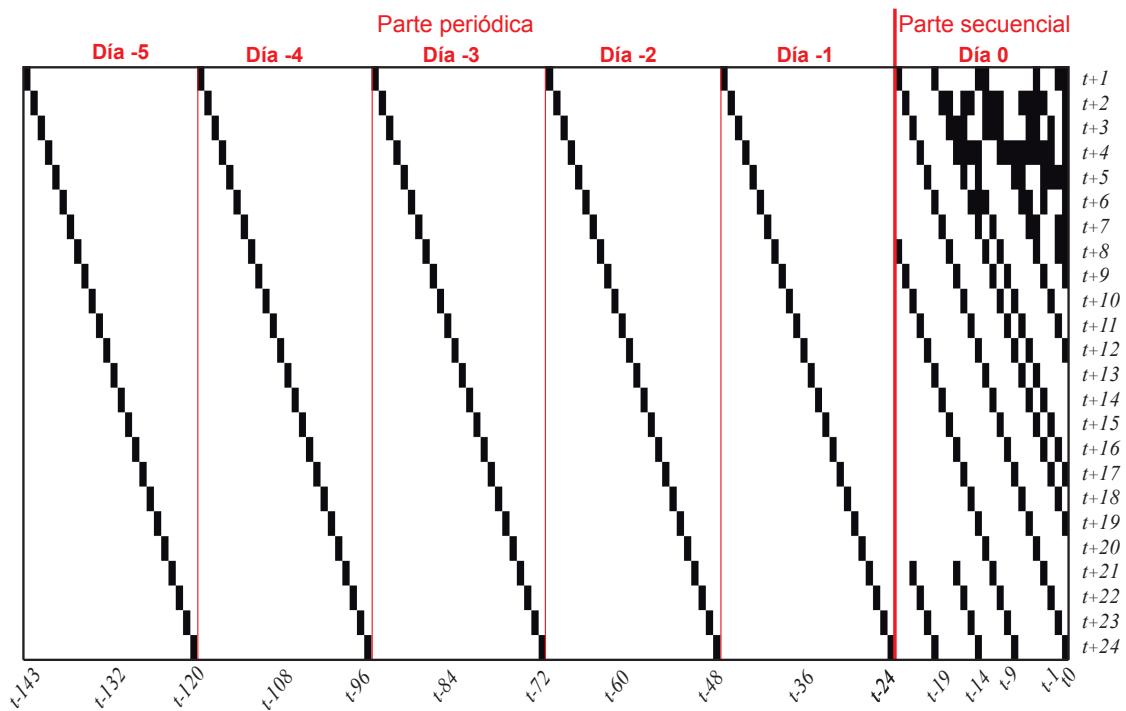


Figura 6.26: Selección del regresor para predicción directa correspondiente al edificio de Filosofía.

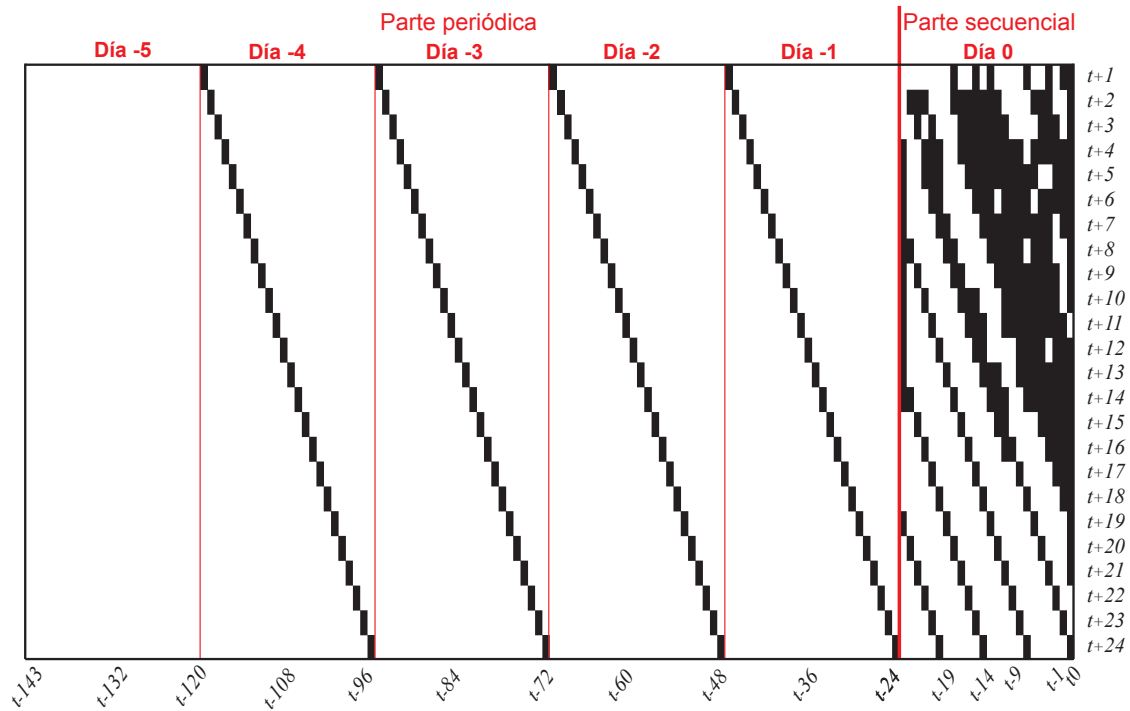


Figura 6.27: Selección del regresor para predicción directa correspondiente al edificio de Veterinaria.

Se puede ver cómo, para cada instante de tiempo, la selección del regresor es completamente independiente y parece aleatoria en parte, lo que coincide con el hecho de que el perfil de potencia no sea dependiente del tiempo.

En el caso de los edificios del *cluster 2*, son edificios muy dependientes de la hora del día, pero se hace la búsqueda de regresor considerando la parte secuencial y la parte periódica. Se puede ver en la figura 6.26 cómo la dependencia es mayormente de la parte periódica; y en la búsqueda de la parte secuencial, cómo existe también una dependencia horaria, al estar siempre relacionada con las horas anteriores.

Por último, para los edificios del *cluster 3*, en los que la dependencia horaria es menor, pero siguen teniendo un perfil marcado por la hora del día, se va a buscar el regresor también en la parte secuencial y en la periódica. Se puede ver en la figura 6.27 que existe una mayor dependencia de la parte secuencial, siendo el regresor mucho más dependiente de la parte secuencial para las horas más cercanas al punto de predicción. Por ello, tanto para este cluster como para el anterior se va a utilizar un regresor para predecir cada instante de tiempo, lo que complica y alarga el entrenamiento de los métodos de predicción.

Predicción directa con KSOM

En la predicción basada en SOM por medio de un método directo, se va a predecir un único instante de tiempo por modelo. Al igual que en el caso MIMO, el regresor también está formado por tres partes (ambiental, regresor y predicción). A diferencia de este, la parte

de predicción estará compuesta por una única muestra, y la parte del regresor dependerá del instante de tiempo a predecir dentro de las siguientes 24 horas, estando formado de distinta manera en función de los resultados obtenidos en la selección del regresor.

El KSOM también se entrena usando una máscara similar, de tal forma que, en el entrenamiento, la neurona ganadora se calcula únicamente con la información de la neurona correspondiente al regresor. De esta forma se logra que el KSOM se entrene en base a los regresores. Dado que en todos los edificios los regresores son diferentes para cada hora del día, es necesario entrenar 24 modelos, uno para cada hora del día.

Los parámetros de entrenamiento son similares al caso del SOM MIMO, pero que además hay que determinar la K de salida del SOM, que es la cantidad de neuronas se utilizan para calcular el modelo autorregresivo con el que realizar la predicción. Por medio de validación cruzada se obtiene que, para edificios como Filosofía, el número de neuronas se encuentra en torno a 20, mientras que para Veterinaria el número de neuronas es de 40 y para Animalario este número asciende a 100. De esto se concluye que cuanto menos dependiente de la hora sea el edificio, mayor es el número de neuronas necesarias en la salida para calcular el modelo. En cuanto al tamaño del SOM, dado que el número de muestras sigue siendo la misma, se utilizan 2000 neuronas y las dimensiones son 50×40 .

Predicción directa con SVR y *Neural Gas*

Este último método agrupa los tipos de días a predecir y asocia un modelo SVR a cada *cluster* para cada hora a predecir de las siguientes 24 horas. Primeramente, con las variables ambientales de la tabla 6.6, se entrena un *Neural Gas*. A estas variables ambientales se añade información de la evolución de la temperatura incorporando la temperatura media de las 24 horas previas \bar{T}_{-24} ya que mejora los resultados en el agrupamiento. Además, también se añade información de la potencia media de las 24 y 48 horas antes (\bar{P}_{-24} y \bar{P}_{-48}) para recoger en los *clusters* que agrupan cada tipo de consumo la evolución de la variable a predecir y no solo en el regresor, lo que ayuda a acotar el *cluster* ganador. El número óptimo de *clusters* se determina en base al índice de *Davies-Bouldin*, obteniéndose que el número óptimo es de 32.

Una vez obtenido el número de *clusters*, se entrenan para cada uno de los grupos 24 modelos SVR, correspondientes cada uno a cada hora en adelante que se va a predecir. El entrenamiento de estos modelos SVR requiere de una gran cantidad de parámetros que hay que sintonizar por medio de validación cruzada. Todos estos parámetros son sintonizados automáticamente por las funciones de la librería, utilizando una combinación de los métodos *Coupled Simulated Annealing* (CSA), para un ajuste inicial de los parámetros, y el método *simplex*, para un ajuste fino de los mismos (Suykens *et al.*, 2002b). Sin embargo, requiere de una gran cantidad de tiempo que complica la ejecución del algoritmo.

6.5.3. Resultados de la predicción a corto plazo

En la tabla 6.7 se muestran los resultados de los índices de error de los métodos de predicción para cada uno de los tres edificios representativos. En las figuras 6.29, 6.31 y 6.33

se muestra el error NRMSE resultante de aplicar los métodos de predicción a corto plazo para cada uno de los edificios representativos de cada grupo para una semana en concreto, no se muestra el CCA ya que el error es muy elevado y distorsiona la escala de la figura.

En todos los casos, el mejor método resultante es el del SVR, lo cual concuerda con los resultados de otros artículos de la bibliografía donde se pone de relieve el buen funcionamiento de esta técnica para la regresión. Al agrupar los instantes de tiempo similares se mejora la precisión de los modelos de regresión ya que solo tiene en consideración regresores con dinámicas similares para la generación del modelo.

El principal problema de este método es que los tiempos de ejecución son muy largos. En comparación con el entrenamiento de un KSOM, el SVR tarda seis horas en entrenarse frente a dos horas que tarda el SOM sobre la misma plataforma. En cuanto al SOM MIMO los tiempos son mucho más cortos, siendo el tiempo en torno a una hora ya que únicamente es necesario entrenar un único modelo.

El método MIMO ofrece un error más constante a lo largo de la ventana de predicción mientras que los directos tienden a aumentar el error conforme la ventana de predicción es mayor, esto es consecuente con el hecho de que los métodos MIMO predicen, en un único paso, todas las muestras futuras, mientras que los métodos directos utilizan una función por cada hora a predecir.

El método que peor rendimiento ofrece de todos los analizados es el método que implementa CCA y AR, lo cual, como se explicó en la sección 3.6, se debe a que la mezcla de tantos métodos para predecir va acumulando una serie de errores que se suman para la predicción final. Además, cada método requiere del uso de la sintonización de una serie de parámetros por lo que el tiempo de entrenamiento también es elevado, por lo que solo resulta útil cuando se trabaja con diferentes variables.

Comparando edificios con el índice NRMSE, el edificio en el que funciona mejor la predicción a corto plazo es el edificio de Filosofía, al igual que ocurría con la predicción a largo plazo. Si se utiliza el índice MAPE, la predicción funciona mejor en Veterinaria, pero como ya se dijo, se debe a que el nivel de consumo es más elevado. Para el edificio de Animalario ninguno de los métodos resulta preciso, dado que el error NRMSE es superior al 10 % y el índice ThU indica que ninguno de los métodos es mucho mejor que realizar una predicción considerando un método simple.

En las figuras 6.28, 6.30 y 6.32 se puede ver la precisión de la predicción de los tres métodos con menor error para una semana concreta para cada edificio cuando el horizonte de predicción es de 24 horas que es la situación más desfavorable. Esta semana mostrada comienza en martes puesto que la predicción se inicia en un lunes. En el caso de Veterinaria y Filosofía, se aprecia cómo las predicciones se adaptan bien a las curvas. Sin embargo, en el caso de Animalario apenas se consigue ajuste, esto demuestra que la predicción para edificios dependientes de la hora puede resultar una tarea sencilla, mientras que para los que no lo son, resulta complejo. En el caso de Veterinaria y Filosofía, el KSOM presenta un error elevado debido a que en los fines de semana (horas entre 100 y 150) acumula cierta inercia, generando una gran desviación en la curva predicha, lo cual empeora los resultados. Para predicción a 24 horas, en este edificio, genera mayor precisión el SOM MIMO.

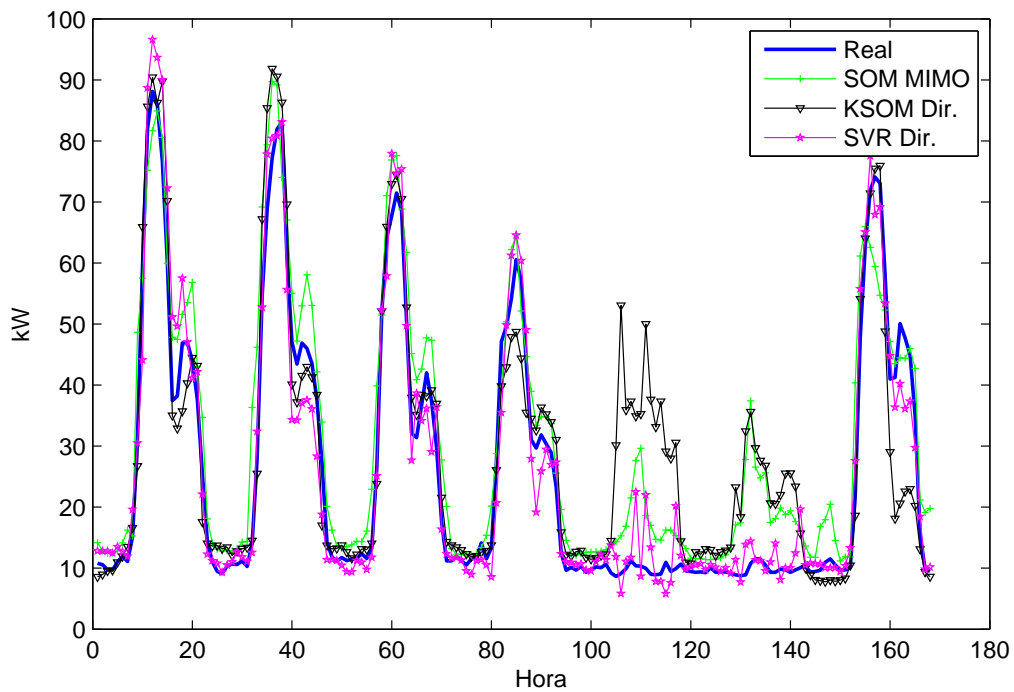


Figura 6.28: Resultados de predicción a corto plazo en Filosofía para el instante $t + 24$.

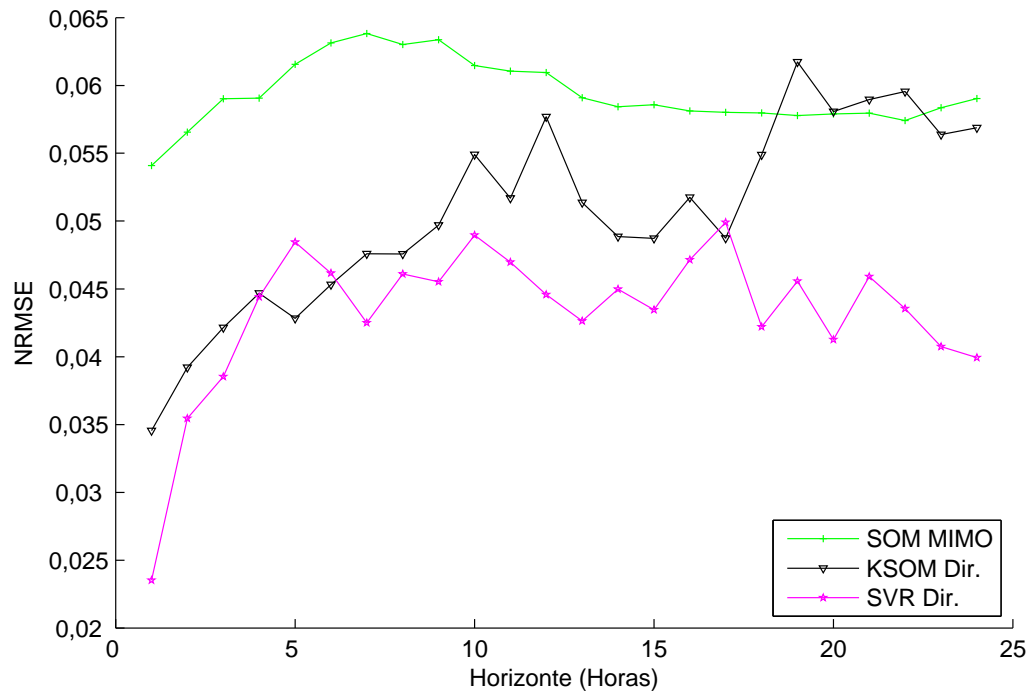


Figura 6.29: NRMSE para los distintos métodos en Filosofía.

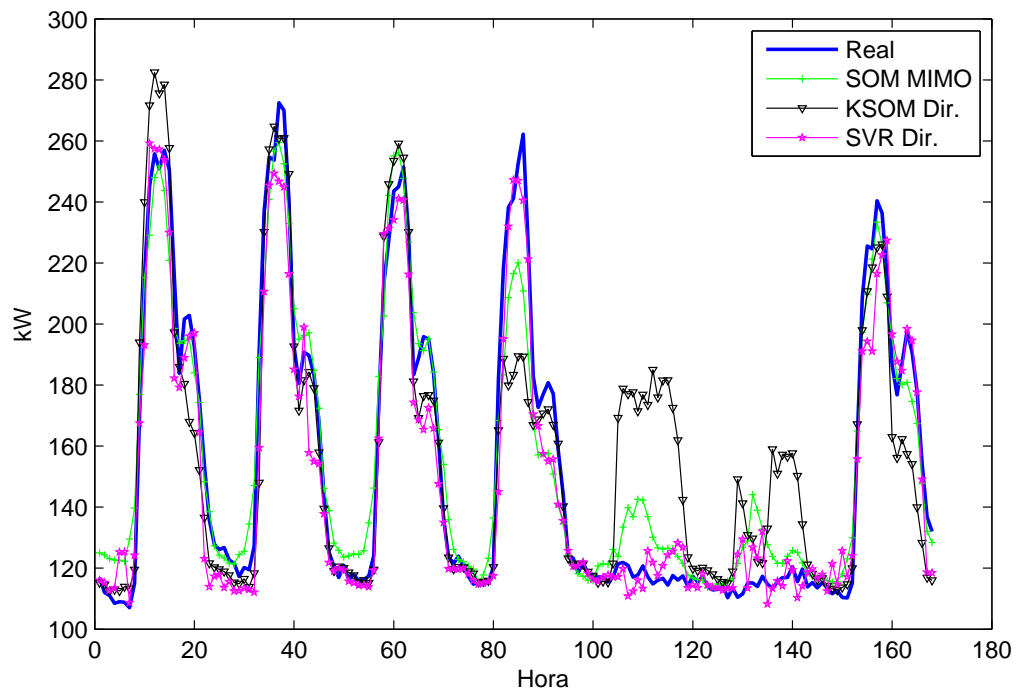


Figura 6.30: Resultados de predicción a corto plazo en Veterinaria para el instante $t + 24$.

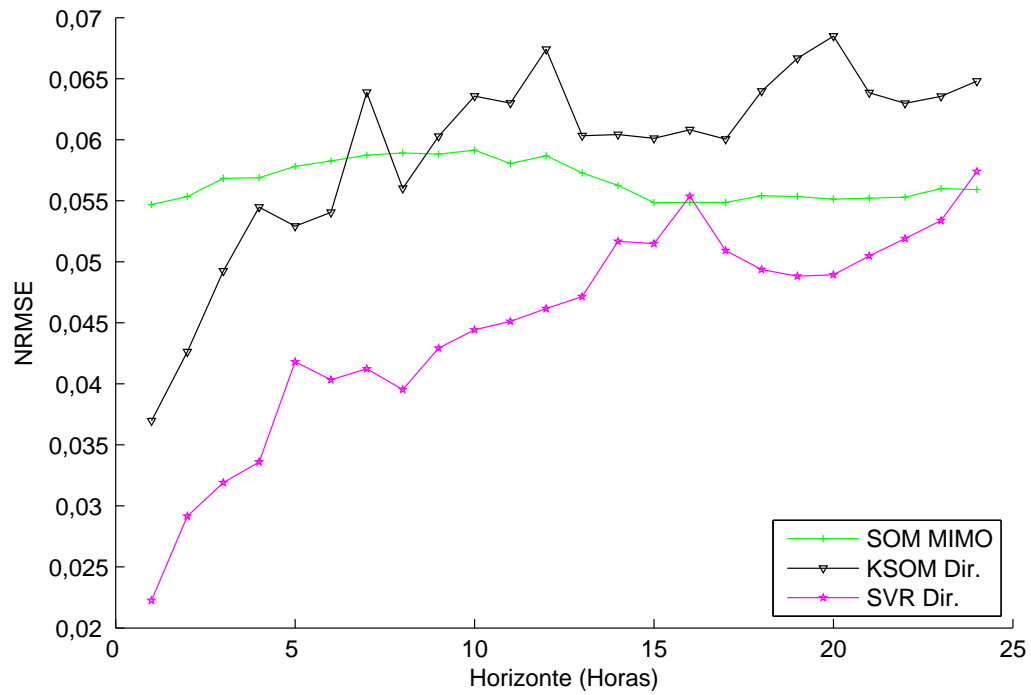


Figura 6.31: NRMSE para los distintos métodos en Veterinaria.

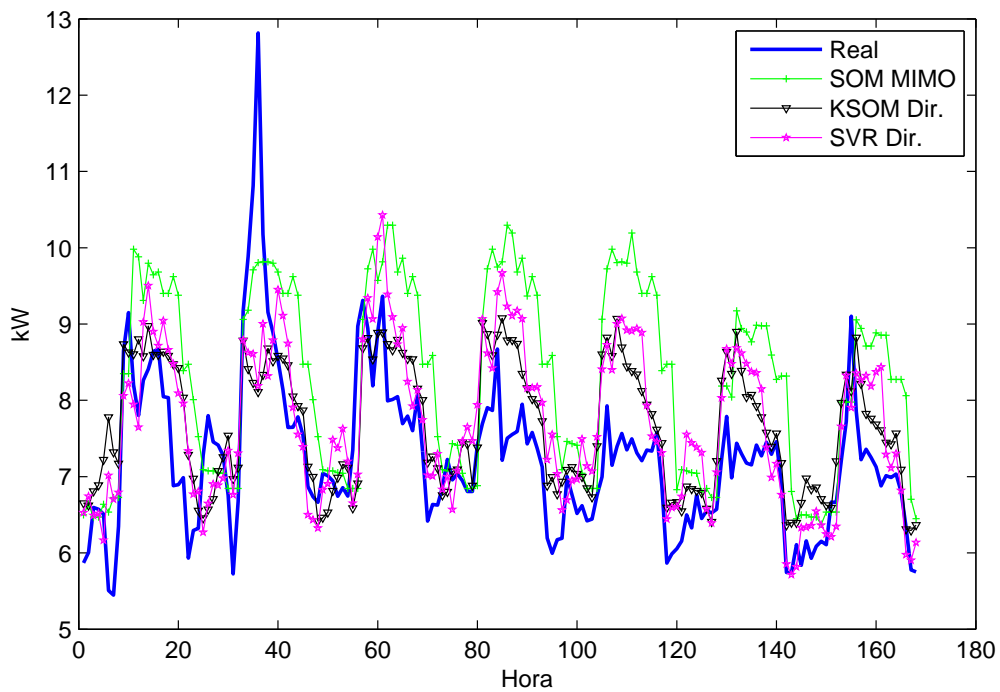


Figura 6.32: Resultados de predicción a corto plazo en Animalario para el instante $t + 24$.

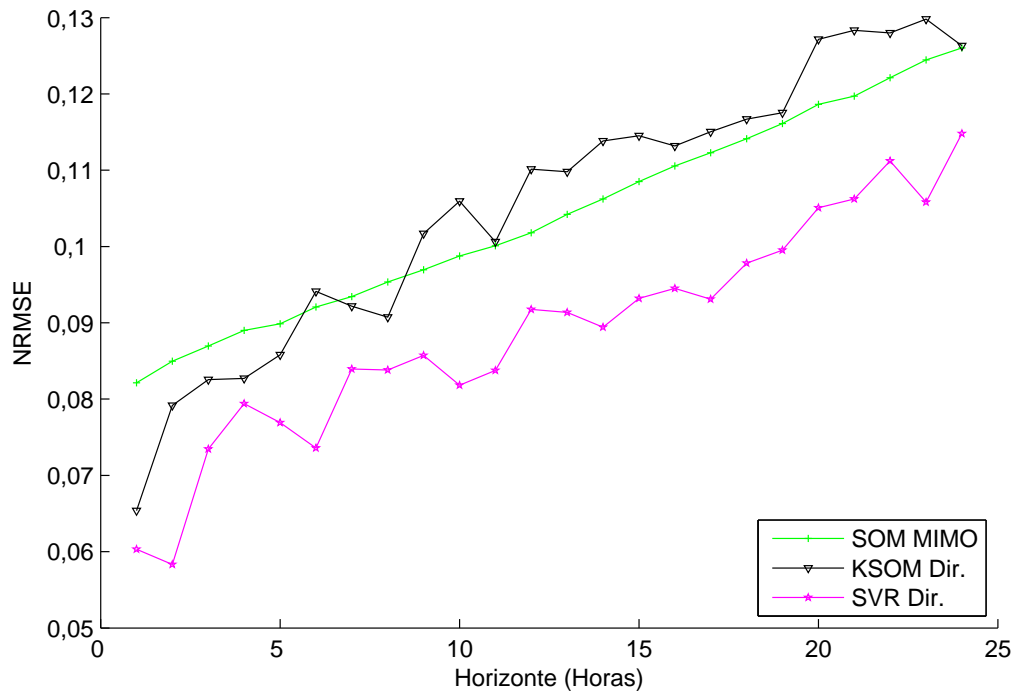


Figura 6.33: NRMSE para los distintos métodos en Animalario.

	Métodos MIMO		Métodos Directos	
	CCA+AR	SOM MIMO	KSOM Dir.	SVR Dir.
Filosofía				
MAPE	26,8213	26,4701	26,5276	17,7506
NRMSE	0,1702	0,0594	0,0506	0,0433
ThU	0,7026	0,4684	0,3958	0,3733
Veterinaria				
MAPE	16,1878	8,4094	7,6362	6,0502
NRMSE	0,2189	0,0566	0,0562	0,0448
ThU	0,8966	0,46224	0,4766	0,3802
Animalario				
MAPE	21,8661	17,6986	15,5634	15,1345
NRMSE	0,1965	0,1039	0,1055	0,0890
ThU	1,0083	0,9537	0,9676	0,8149

Tabla 6.7: Índices de predicción para la medida del error de predicción a corto plazo.

6.6. Visualización de facturas

La facturación de grandes edificios no se hace solamente en base al consumo de energía activa, como en el caso de consumos domésticos, sino que se considera también la energía reactiva y el pico de potencia demandada. Por ello, la visualización de la evolución de la facturación no se puede hacer en base únicamente a gráficos de barras, como suelen hacer las compañías eléctricas, ya que se omite información que influye de una manera importante en el coste de facturación. Dado que son varias variables y que la Universidad de León tiene varios puntos de facturación, la visualización por medio de gráficos de barras puede complicar tanto la extracción de conclusiones de los datos como la comparación de los puntos de facturación entre sí.

Para solucionar este problema, se plantea utilizar herramientas de reducción de la dimensionalidad para proyectar los datos de las facturas de todos los puntos de facturación en un mapa de dos dimensiones. Estos mapas, similares a los de comparación de perfiles, permiten proyectar en un único punto bidimensional los datos relativos a una factura. Dado que la proyección es discreta, se utilizan mapas continuos para poder obtener visualizaciones que muestren no solo los datos, sino también las fronteras a partir de las cuales las facturas se ven penalizadas. La proyección de los puntos con estas fronteras permite ver, de forma rápida, las facturas que están sometidas a penalizaciones.

En la tabla 6.8 se muestran las variables que se han utilizado para realizar la proyección de las facturas. Normalmente, la compañía presenta gráficas de evolución únicamente del consumo de energía activa en las facturas, por lo que aparte de estas variables se han añadido todas aquellas variables que influyen en la factura y que además están sujetas a penalizaciones. Otras variables, como el precio de la energía, aunque también marcan el importe, son parámetros sobre los cuales no se puede actuar, y por tanto no aportan

Identificador	Variable	Tipo de ponderación
DP_p	Desviación maxímetro punta	1
DP_{ll}	Desviación maxímetro llano	1
DP_v	Desviación maxímetro valle	1
Ea_p	Energía activa punta	2
Ea_{ll}	Energía activa llano	2
Ea_v	Energía activa valle	2
Er_p	Energía reactiva punta	3
Er_{ll}	Energía reactiva llano	3
Er_v	Energía reactiva valle	3
$\cos \phi_p$	$\cos \phi$ punta	4
$\cos \phi_{ll}$	$\cos \phi$ llano	4

Tabla 6.8: Variables de las facturas empleadas para la reducción de la dimensión.

información y no influyen en la proyección. El objetivo fundamental es la visualización del consumo y las penalizaciones, no tanto de ver la evolución del importe de las facturas.

La técnica utilizada para reducir la dimensión en este caso ha sido el SNE, ya que proporciona buenos resultados visuales y es la que menos parámetros necesita para el entrenamiento. Un motivo adicional para el uso de esta técnica es porque la continuidad que presenta esta técnica es alta, lo que permitirá una buena generación de mapas continuos.

Para generar la proyección, se crean vectores con las variables de la tabla 6.8 con el fin de construir el espacio de entrada. Algunas de estas variables son ponderadas para aumentar la distancia en el espacio de entrada y así lograr que se proyecten a mayor distancia en el espacio de salida, destacando del resto. La ponderación se hace acorde a la penalización que sufre esa variable en la factura. El tipo de ponderación recogido en la tabla indican el índice de la siguiente lista, que es realizada para cada variable:

1. **Desviación de la potencia contratada.** Potencia máxima consumida (P_m) – Potencia contratada (P_c). La compañía factura la potencia máxima consumida (P_f) de acuerdo a la siguiente ecuación.

$$P_f = \begin{cases} 0,85P_c & \text{si } P_m < 0,85P_c \\ P_c & \text{si } 0,85P_c \leq P_m \leq 1,05P_c \\ P_m + 2(P_m - 1,05P_c) & \text{si } P_m > 1,05P_c \end{cases} \quad (6.9)$$

La ponderación de los términos de desviación de la potencia se hace de la misma forma que lo realiza la compañía.

2. **Energía activa.** No se pondera, ya que la compañía no penaliza este término.
3. **Energía reactiva.** No se pondera, puesto que la penalización se hace sobre el $\cos \phi$, que es lo que realmente interesa conocer, mientras que conocer la energía reactiva es interesante para poder comparar los consumos de reactiva de los puntos de facturación entre sí.

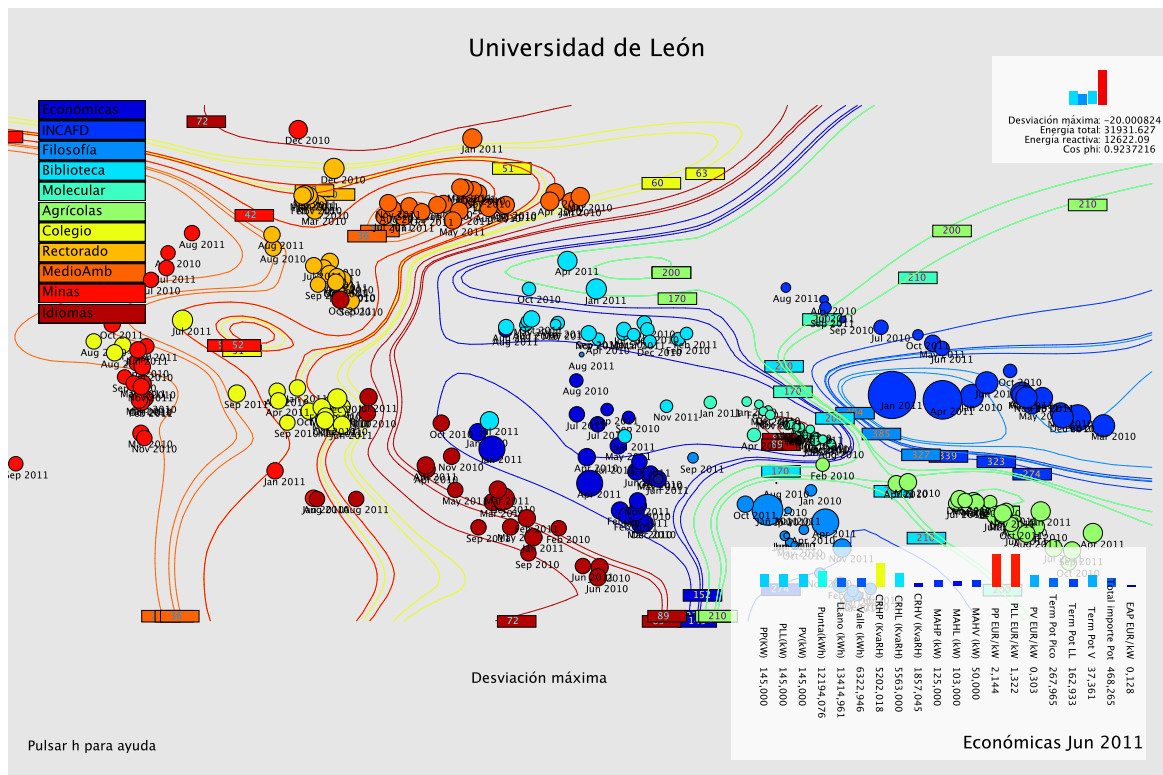


Figura 6.34: Herramienta en *Processing* para la visualización de las facturas.

4. $\cos \phi$. El término utilizado es $1 - \cos \phi$ con el fin de que valores bajos de $\cos \phi$ permita generar valores altos en la proyección.

- $\cos \phi > 0,95$. No se pondera porque no existe penalización. No se considera en ningún momento la posibilidad de que exista un $\cos \phi$ capacitivo.
- $0,95 > \cos \phi > 0,8$. Se pondera por un factor de 2.
- $0,8 > \cos \phi$. Se pondera por un factor de 5 ya que la penalización es elevada.

La elección del factor de ponderación se hace en base a lo que se incrementa el precio con el que se grava la potencia reactiva que se ha consumido cuando se ha excedido el $\cos \phi$.

Una vez realizada la proyección de los datos, se entrena el mapa continuo mediante una red GRNN con una $\sigma = 2$, con el fin de visualizar, por medio de una representación de curvas de nivel, los límites que suponen una penalización por parte de la compañía.

El principal objetivo de esta técnica es el desarrollo de una herramienta que permita a un usuario obtener información de forma rápida y sencilla acerca de los meses que sufren una mayor desviación en la factura y que, por tanto, son objeto de una mayor penalización. Para ello se ha programado una aplicación en *Processing* en la que se visualiza simultáneamente la proyección de los puntos de facturación durante los años 2010 y 2011.

La herramienta de *Processing* (ver figura 6.34) permite la selección de los puntos de facturación a visualizar. De esta forma, se puede realizar un análisis de un único punto de

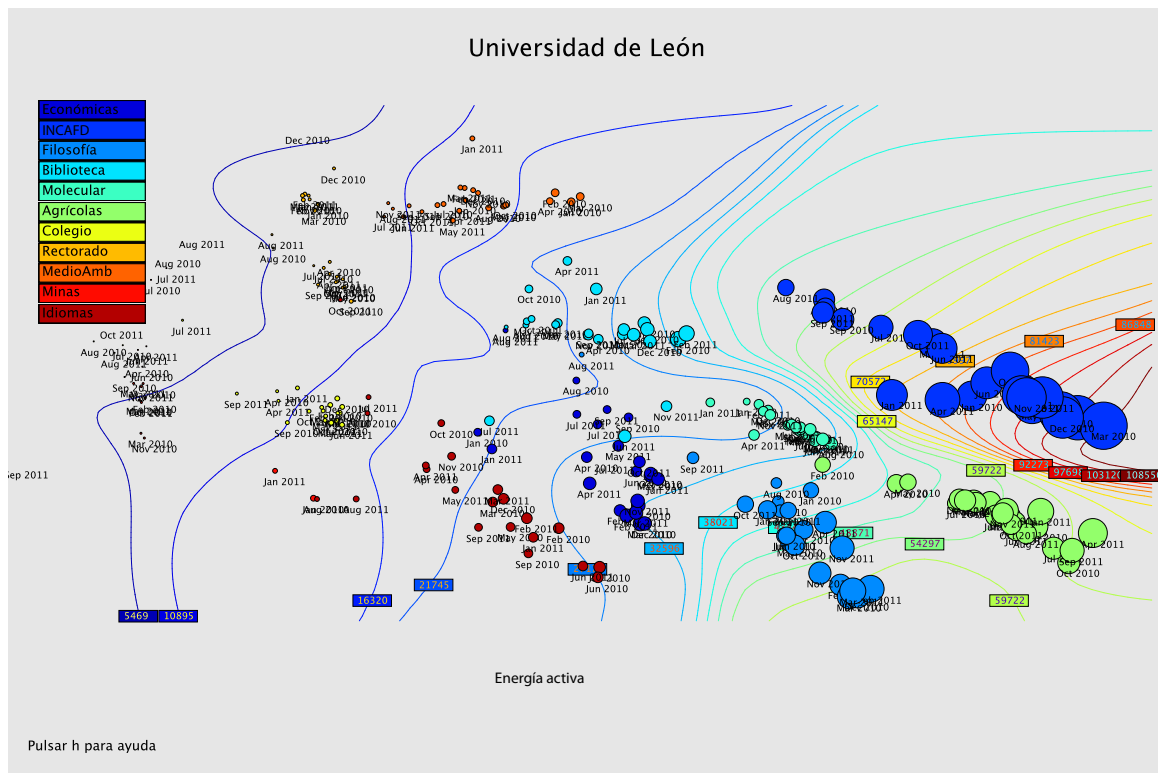


Figura 6.35: Visualización de la energía activa consumida en cada factura.

facturación a lo largo de estos dos años y ver los meses con desvíos más claros. También se pueden proyectar todos los puntos de facturación simultáneamente con el fin de poder compararlos entre sí y ver qué facturas tienen un mayor importe, debido bien sea a un mayor consumo o a una mayor penalización. Los colores de los puntos proyectados se corresponden con cada edificio y se muestran en la lista cuando se seleccionan.

La herramienta permite la selección de cada factura para ver en mayor detalle todos los datos de dicha factura, como se ve en la gráfica de la parte inferior derecha de la figura 6.34, apareciendo en la parte superior el detalle de los valores de los términos que suponen una penalización en la factura. En el caso de estas gráficas el color es acorde a los valores máximos registrados en las facturas, de tal manera, un color rojo indica un valor cercano al máximo registrado y un color verde es un valor cercano al mínimo.

También se puede mostrar una gráfica con la evolución de la potencia máxima diaria a lo largo del periodo de facturación, con el fin de poder determinar en que día se ha producido la mayor penalización. Esto será útil durante un estudio posterior del comportamiento del edificio ya que se podrán analizar las condiciones que han propiciado esa situación.

En la herramienta, el tamaño de punto de cada factura es proporcional a una de las siguientes variables:

- **Suma de las energías activas consumidas** ($E_{ap} + E_{aul} + E_{av}$). La figura 6.35 muestra los consumos de las facturas para todos los puntos de facturación con un mapa de curvas de nivel que indican la suma de energía consumida, siendo azul para

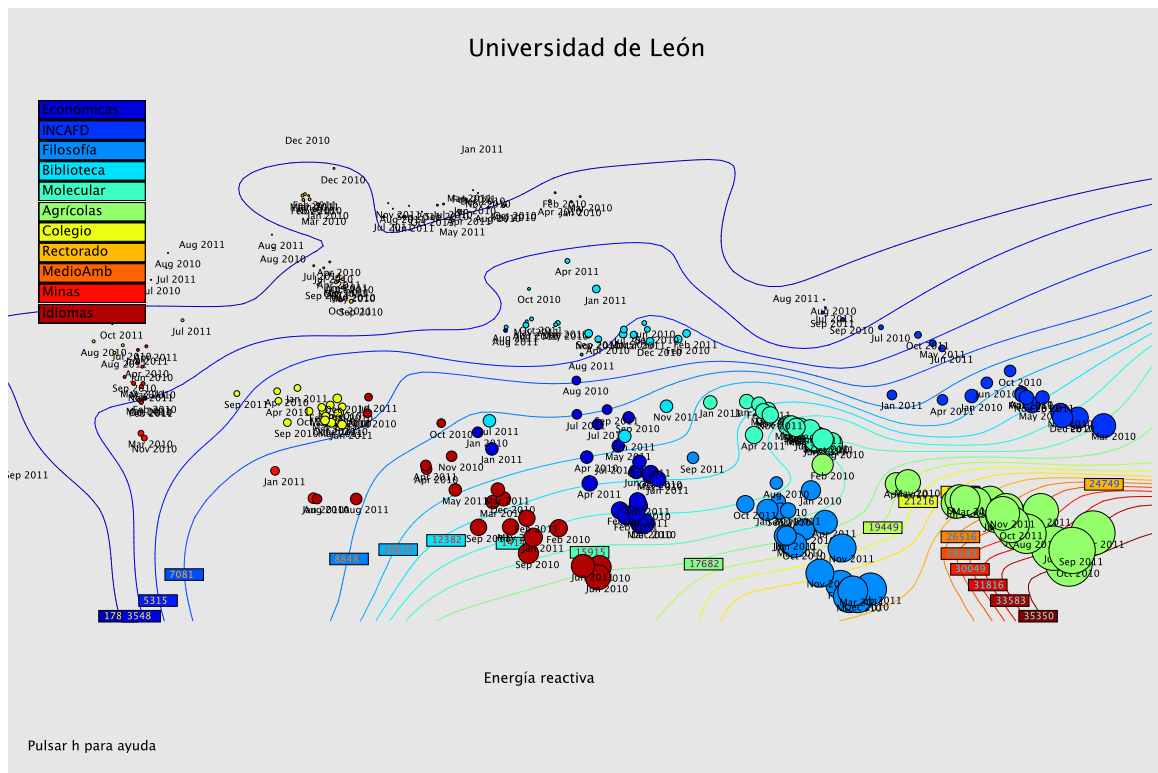


Figura 6.36: Visualización de la energía reactiva consumida en cada factura.

los de menor consumo (izquierda del mapa) y rojo para los que presentan consumos mayores (derecha del mapa). Se puede ver que, como es previsible, dentro de cada punto de facturación, los menores consumos se dan siempre en los meses de verano y los mayores, en los de invierno. Haciendo una comparación entre los puntos, se puede apreciar cómo los puntos de facturación que aglutinan mayor número de edificios (ver tabla 4.1), como es el caso de INCAFD, tienen mayores consumos de energía que los que están constituidos por un solo edificio, como es el caso de los puntos situados a la izquierda del mapa.

La visualización de la energía activa da idea de la actividad de los puntos de facturación y del consumo de los mismos. Esta variable permite ver los periodos de mayor consumo. Como es lógico, los mayores consumos se producen en los meses lectivos. Este comportamiento es más evidente en los puntos de facturación que aglutinan edificios docentes, como INCAFD o Económicas, mientras que en puntos de facturación de edificios de investigación, como Medio Ambiente, el consumo es más constante.

- **Suma de las energías reactivas consumidas** ($Er_p + Er_u + Er_v$). La figura 6.36 muestra los consumos de energía reactiva en todos los puntos. Al igual que en la figura anterior, se muestra un mapa de con curvas de nivel para determinar las zonas de mayor (rojo) y menor (azul) consumo. Dado que algunos de los edificios poseen sistemas de baterías de condensadores para compensar la energía reactiva, los puntos de facturación de mayor consumo de energía activa no coinciden con los de mayor energía reactiva. El punto con mayor consumo de energía reactiva en este caso es

el de Agrícolas. En algunos casos las fechas de verano no coinciden necesariamente con un menor consumo de energía reactiva, lo que indica la presencia de equipos de climatización, como es el caso de Biblioteca, o edificios de investigación cuyo consumo de reactiva es más o menos constante.

- **Mínimo de los factores de potencia** ($\min(\cos \phi_p, \cos \phi_u)$). Para el caso del $\cos \phi$ solo se visualiza el menor de los $\cos \phi$, ya que la compañía eléctrica realiza la penalización en función del menor de ellos. En cuanto a las curvas de nivel, solo se muestran dos curvas que son las correspondientes a los valores de un $\cos \phi$ de 0,95 y 0,8, ya que son los que marcan los umbrales de penalización. De un solo vistazo a la figura 6.37 se aprecia claramente cuáles son los puntos de facturación más penalizados y cuáles los menos.

En el caso de INCAFD, Biblioteca y Medio Ambiente se ve cómo el $\cos \phi$ es casi siempre superior a 0,95, por lo que no sufren penalización alguna. Esto se debe a que algunos de los edificios contenidos en estos tres puntos de facturación tienen instaladas baterías de condensadores. Sin embargo, la propia visualización nos muestra que casos como el de Biblioteca están mal dimensionados, ya que hay algún mes en el que el $\cos \phi$ llega a ser penalizado. El resto de puntos sufren siempre algo de penalización, se puede apreciar como Agrícolas, que es el que mayor energía reactiva consume, está en esta franja de penalización. Según este análisis, Minas es el punto de facturación que habría que corregir con más urgencia, puesto que existen meses en los que su $\cos \phi$ baja del umbral de 0,8.

- **Máximo de las desviaciones de potencia demandada** ($\max(DP_p, DP_u, DP_v)$). En la figura 6.38, para facilitar la visualización, se muestran solo tres puntos de facturación con sus curvas de potencia máxima demandada. En este caso, las curvas de nivel tienen el mismo color que el edificio al que representan y las etiquetas numéricas muestran el valor correspondiente a esa curva. Dado que la compañía eléctrica penaliza los máximos de potencia que superen en un 5 % o no lleguen al 85 % las curvas de nivel determinan una franja en la que debe estar el máximo demandado para que no sea objeto de penalización. Dado que el máximo valor de potencia es un término que se puede pactar con la compañía, esta visualización permite ver qué edificios tienen un máximo contratado correcto o en cuáles se debería cambiar.

De los casos representados, se puede ver claramente cómo en Económicas el máximo no sería necesario modificarlo ya que la mayoría de los puntos caen dentro de la franja, salvo por los meses de verano (en general el consumo cae en estos meses, por lo que no sería necesario mejorarlo). Sin embargo, en el caso del punto de facturación correspondiente al Molecular se ve cómo el máximo nunca llega siquiera al mínimo de potencia contratada, por lo que sería necesario reducir el máximo contratado para disminuir el importe de la factura.

Esta herramienta permite cambiar las curvas de desviación para mostrar las curvas que se corresponderían con una potencia contratada óptima, la cual minimizaría las penalizaciones por desviaciones en la potencia contratada. Estas curvas se calculan

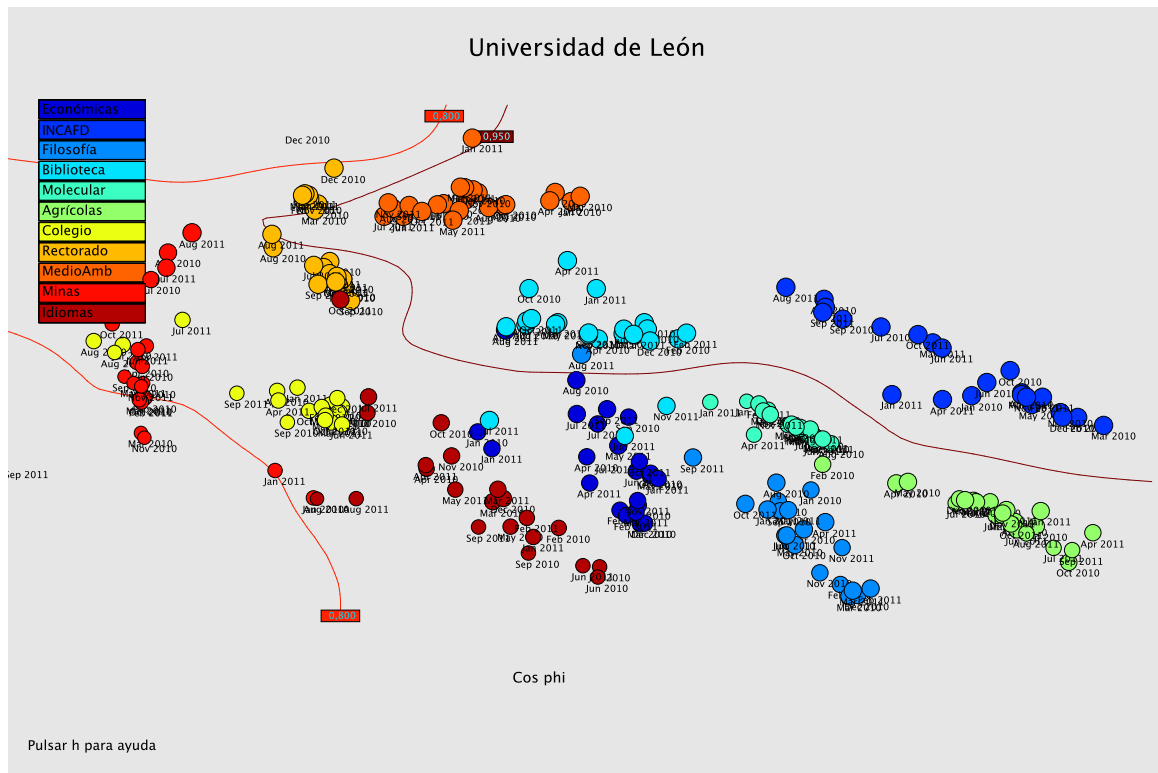
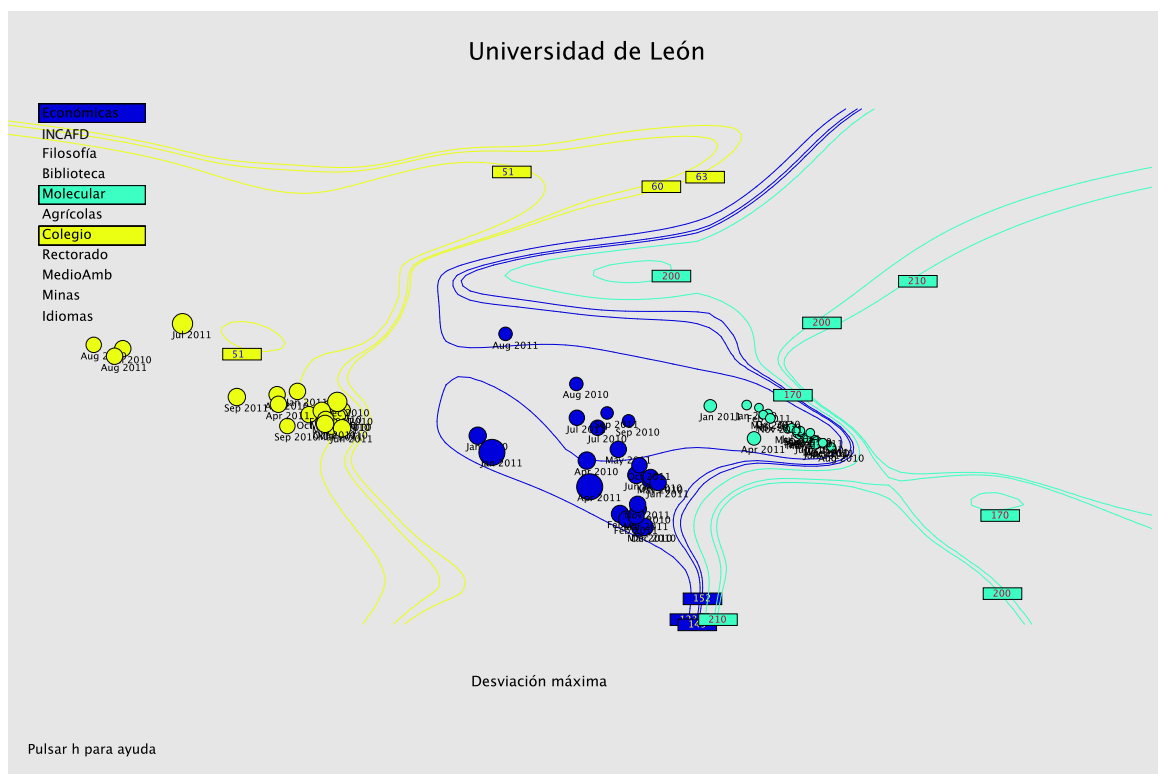
Figura 6.37: Visualización del $\cos \phi$ en cada factura.

Figura 6.38: Visualización de la potencia pico marcada por el máxímetro.

Punto	P contratada (kW)	P óptima (kW)	Ahorro (€)
Económicas	380	323	1011,51
INCAFD	385	249	3422,76
Filosofía	200	134	1665,77
Biblioteca	145	137	91,92
Molecular	200	68	4441,75
Agrícolas	200	191	91,48
Colegio	60	45	200,4
Rectorado	50	44	42,6
Medio Ambiente	35	41	46,8
Minas	50	28	378,83
Idiomas	35	58	428,56

Tabla 6.9: Ahorro anual por la corrección de la potencia contratada en las facturas de la Universidad.

minimizando la función de cálculo del término de potencia, expresado en la ecuación 6.9 con todas las facturas que se han utilizado para la proyección para ese punto de facturación. En la tabla 6.9 se recogen los ahorros anuales producidos gracias a la corrección de la potencia contratada. En esta tabla se puede ver cómo puntos de facturación como el de Molecular tienen una elevada desviación de potencia que supone un ahorro de más de 4000 €. Se puede ver cómo solamente los puntos de facturación de Agrícolas y Rectorado son los únicos en los que la potencia contratada es acorde a la demandada. El edificio de Medio Ambiente tiene una desviación también pequeña, pero que dado que es una desviación por exceso y están mucho más penalizadas, es conveniente corregirlo.

Por medio del análisis de las facturas utilizando esta herramienta, se ha podido ver el comportamiento de todos los puntos de facturación de la Universidad de León. Puntos como Molecular sufren elevadas penalizaciones debido fundamentalmente a su elevada desviación de la potencia contratada, mientras que sufren poca penalización debido al $\cos \phi$, por lo que su corrección resulta fácil, ya que solo hay que renegociar el contrato con la compañía. Puntos como Minas, aunque presentan poca desviación, sí sufren una elevada penalización por culpa de un $\cos \phi$ muy bajo, por lo que su corrección necesita de la instalación de baterías de condensadores, lo cual requiere una inversión previa. En cuanto a puntos como Agrícolas e INCAFD, sufren pocas penalizaciones y el coste de las facturas se debe principalmente a la gran cantidad de energía consumida.

Conclusiones y líneas futuras

En este capítulo se exponen las conclusiones obtenidas en este trabajo. También se enumeran las aportaciones que han resultado de esta tesis, clasificándose en científicas y técnicas. Finalmente, se mencionan las posibles líneas futuras de trabajo que se desprenden de esta tesis.

7.1. Conclusiones

A lo largo de este trabajo se ha propuesto una metodología basada en técnicas de minería de datos y reducción de la dimensionalidad para la supervisión y predicción de consumos energéticos en edificios públicos de uso docente. Esta metodología ha sido probada con los consumos eléctricos de los edificios que conforman el campus de la Universidad de León. Se ha mejorado el sistema de supervisión ya existente, haciéndolo más escalable y fácilmente configurable, con el fin de que no solo sirva para la adquisición de datos eléctricos sino para cualquier otro tipo de datos. A partir de los datos obtenidos por medio de este sistema, se han aplicado técnicas de reducción de la dimensionalidad con el fin de lograr proyecciones interactivas sobre las que comparar los consumos de los distintos edificios entre sí y poder visualizar también la dependencia de un conjunto de variables ambientales que tiene el consumo de los edificios. A partir de este análisis, se han podido escoger una serie de variables ambientales relevantes y se han podido agrupar los edificios entre sí de acuerdo a su comportamiento dinámico. Con estos datos se han creado los regresores para cada tipo de edificio y se han aplicado diversas técnicas de predicción a corto y medio plazo con el fin de poder estimar el consumo futuro de estos edificios.

Las herramientas de visualización ayudan a comprender el comportamiento de los edificios y a identificar fácilmente aquellos consumos que salen fuera del rango normal. Además, se han aplicado estas técnicas a otro tipo de datos, como son los correspondientes a las facturas, con el objetivo de lograr una interpretación y análisis de una forma más intuitiva.

A la vista de los resultados, las conclusiones que se pueden extraer de esta tesis son las siguientes:

- Se ha mejorado el sistema de supervisión utilizando una estructura de triple capa, que permite la adquisición, almacenamiento y visualización de los datos. Esta estructura

ha sido planteada de forma modular y escalable por medio de una capa intermedia con tareas independientes, aunque utilicen el mismo hardware. La estructura ha sido puesta en marcha con los datos eléctricos procedentes de los medidores instalados en los edificios; sin embargo, está preparada para ser utilizada con otro tipo de datos haciendo pequeñas modificaciones.

- Se han estudiado las correlaciones de una serie de variables ambientales con la evolución del consumo energético en los edificios. Se ha podido determinar cómo la correlación de estas variables depende, en enorme medida, de la actividad principal del edificio. Así, los edificios cuya actividad principal es la docencia tienen una mayor correlación, mientras que los edificios destinados a investigación presentan una correlación mucho menor, incluso inexistente. Asimismo, se ha estudiado la dependencia horaria del consumo en los edificios obteniéndose conclusiones análogas a las de las variables ambientales.
- Se ha realizado un análisis de diferentes técnicas de reducción de la dimensionalidad y se ha comprobado su eficacia para la generación de visualizaciones con los datos eléctricos. Realizando un análisis visual de todas las técnicas y apoyando este estudio con el empleo de métricas que miden la eficacia de una proyección, se ha determinado cuáles son las mejores técnicas a utilizar en el caso de las curvas de consumo eléctrico, siendo las mejores el SNE y el t-SNE.
- Se han propuesto dos metodologías para el análisis visual de los datos por medio de técnicas de minería de datos. Por un lado, se ha propuesto el uso de técnicas de reducción de la dimensionalidad, como el SNE y el t-SNE, junto con el uso de técnicas como el GRNN para la generación de mapas continuos en los que ver, por medio de una proyección 2D, la distribución de las curvas de consumo y las variables ambientales. Por otro lado, se ha propuesto la creación de modelos por medio del uso de técnicas como el SOM. Estos modelos permiten relacionar las variables ambientales con las curvas de consumo específicas para un día dado.
- Se ha propuesto una herramienta de visualización, basada en *Processing*, a partir de las mapas continuos, para la comparación de los edificios del campus de la Universidad de León en base, tanto al nivel de consumo (cantidad de energía diaria), como a la dinámica de dicho consumo (forma de la curva de consumo). Esta comparación permite agrupar edificios en función de su actividad, ya que tienden a proyectarse de la misma forma, y también determinar en qué edificios tendría mayor influencia una corrección para reducir el consumo.
- Se ha propuesto una nueva aplicación de los mapas de diferencias del SOM, al utilizarlos para comparar modelos de consumo entre edificios. Utilizando como base común las variables ambientales, es posible comparar neuronas similares en los dos modelos, mientras que el análisis de las diferencias que se producen entre los modelos hace posible obtener conclusiones acerca de la similitud en la actividad de los dos edificios.

- Se ha propuesto una herramienta de predicción a largo plazo que permite hacer estimaciones de los niveles de consumos de los edificios utilizando los modelos obtenidos con el SOM. Estas estimaciones se obtienen en base a una predicción conocida de los valores de las variables ambientales. Obteniendo diversas predicciones meteorológicas para ese día, se puede obtener un límite máximo y uno mínimo de predicción en el que va a oscilar el consumo estimado para ese día. Esta estimación es útil para la planificación del consumo de un edificio. Además se ha desarrollado una herramienta en *Processing* para facilitar el cálculo de estos límites de consumo para todos los edificios.
- Se ha demostrado la necesidad de procesar previamente los datos con el fin de construir los regresores más adecuados para la predicción. Para ello, se ha llevado a cabo un análisis por medio de técnicas de selección de tipo de filtrado, como el k -NN, para determinar la dependencia del valor predicho con respecto a los valores pasados de la curva de consumo. Este análisis, junto con la información aportada por el análisis de la correlación con la hora del día, permite añadir información relevante acerca del consumo pasado del edificio.
- Se ha llevado a cabo una comparación entre diversos métodos de predicción a corto plazo, basados en técnicas de reducción de la dimensionalidad, para predecir las curvas de consumo a 24 horas. Se han probado métodos MIMO y Directos, métodos basados en el SOM, *clustering* y *manifold learning*. Todos los métodos implementados, en función de la actividad del edificio, han dado diferentes precisiones en la estimación del consumo, siendo el mejor método para todos los casos la combinación del *Neural Gas* y el SVR. Este método, que consta de dos partes, aporta la mejor precisión para todos los tipos de edificios. Para edificios con una elevada dependencia del consumo con la hora del día, se logran resultados de elevada precisión. Sin embargo, para edificios de carácter no docente, que tienen muy poca dependencia horaria, cualquier método se torna ineficaz al no ser capaz de mejorar el resultado de una predicción simple.
- Se ha aprovechado el conocimiento extraído en la generación de mapas continuos para aplicarlos en la creación de proyecciones que permitan la visualización de otro tipo de datos. Para ello, se ha utilizado la información de las facturas de la compañía suministradora y se ha creado una visualización continua de los datos de estas facturas utilizando el SNE. Esta visualización no solo aporta información de los datos de la factura, sino también proporciona información de cuál debe ser el máximo de potencia contratada para lograr la menor penalización. También proporciona información sobre los umbrales de $\cos \phi$ entre los que se mueve el edificio.

7.2. Aportaciones

Las contribuciones de esta tesis se pueden clasificar en dos grupos: las aportaciones en el campo de la ciencia e investigación y las aportaciones técnicas en el campo de la ingeniería de sistemas de medida y supervisión de la energía eléctrica.

Aportaciones científicas. Como aportaciones científicas podemos citar las siguientes:

- Se ha definido un sistema de medida y supervisión de la energía eléctrica para todos los edificios de la Universidad de León. Dicho sistema se basa en una arquitectura de red compleja de triple capa. Este trabajo ha sido presentado en las *XXXII Jornadas de Automática*, JA 2011 (Fuertes *et al.*, 2011).
- Se ha comprobado cómo los métodos basados en minería de datos ofrecen un gran potencial en el análisis del consumo de energía eléctrica en edificios públicos. Este trabajo fue presentado en el *18th World Congress of the International Federation of Automatic Control*, IFAC 2011 (Domínguez *et al.*, 2011).
- Se ha determinado el mejor método de reducción de la dimensionalidad para la visualización de las curvas diarias de consumo y su dependencia de las variables ambientales. Además, se han comparado varios edificios entre sí para ver la influencia de los patrones de consumo en uno y otro edificio. Este trabajo ha sido presentado en el *13th International Conference on Engineering Applications of Neural Networks*, EANN 2012 (Morán *et al.*, 2012c).
- Se ha desarrollado un herramienta en *Processing* que permite la comparación de los edificios entre sí con el fin de poder agruparlos dependiendo de su consumo. Este trabajo se presentó en el *VIII Simposio CEA de Control Inteligente* (Morán *et al.*, 2012a).
- Se ha desarrollado una herramienta avanzada para la estimación del consumo de los edificios a largo plazo, en base al conocimiento previo de las variables ambientales. Este trabajo ha sido presentado en el *International Conference on Renewable Energies and Power Quality*, ICREPQ 2012 (Domínguez *et al.*, 2012a).
- Las herramientas de comparación de los edificios, las de predicción a largo plazo y la arquitectura de red de toma de datos de energía eléctrica también se han enviado para su publicación en la revista *Energy and Buildings*.
- Se ha hecho un análisis de los datos con el fin de determinar cuál es el regresor óptimo a la hora de predecir. Para ello se han seguido una serie de pasos que han ido desde la selección de variables hasta la aplicación de técnicas de búsqueda. Este trabajo ha sido aceptado en el *11th International Symposium on Intelligent Data Analysis*, IDA 2012 (Morán *et al.*, 2012d).
- Algunas de las técnicas de predicción basadas en reducción de la dimensionalidad han sido aplicadas con el fin de realizar predicciones MIMO de series temporales de forma efectiva. Este trabajo fue presentado en el congreso PHYSCON 2011 y ha sido aceptado para su publicación en la revista *International Journal of Modern Physics B* (Morán *et al.*, 2011).
- La herramienta de facturación permite la visualización de forma más sencilla de toda la cantidad de datos presente en una factura y detectar los días en los que se supera el

máximo contratado. Este trabajo ha sido presentado en el *10th Portuguese Conference on Automatic Control*, CONTROLO'12 (Morán *et al.*, 2012b).

Aportaciones técnicas. Las principales aportaciones técnicas de ingeniería para la Universidad de León son las siguientes:

- Se ha mejorado la plataforma para la adquisición, almacenamiento y procesamiento de los datos eléctricos. Esta plataforma implementa un sistema de seguridad que evita un acceso por parte de usuarios externos a la red interna de medida. Además, esta plataforma flexible se está adaptando mediante pequeños cambios para el análisis de otras variables energéticas, como es el caso de las calderas de la Universidad.
- Se han desarrollado tres herramientas interactivas utilizando *Processing* para la comparación de perfiles, la predicción de consumo y para la comparación de facturas. Estas herramientas pueden ser utilizadas por personal no cualificado para realizar un análisis de datos de los consumos de la Universidad de León.
- El análisis de los perfiles, junto con la generación de modelos para la estimación de consumos, ha permitido determinar bajo qué situaciones el consumo de los edificios es mayor y analizar por qué se producen estas situaciones. Análisis más detallados de los días de elevados consumos, como en el caso de Veterinaria, han permitido ver picos de consumo debido al arranque de equipos que podían ser pospuestos y, por tanto, reducir así el riesgo de que se sobrepase el máximo contratado.
- La comparación entre todos los edificios permite determinar cuáles tienen un consumo similar con el fin de poder agruparlos entre sí a la hora de determinar los puntos de facturación óptimos en un futuro.
- Se han detectado mediante los mapas continuos días de consumo anómalo debido a una mala programación de equipos de climatización que estaban encendidos en días en los que el edificio no tenía ninguna actividad.
- El análisis por medio de la herramienta de *Processing* de las facturas ha permitido ver qué edificios no tenían una potencia máxima óptima contratada y ha permitido corregir esta deficiencia, consiguiendo así un ahorro en las facturas al reducir la penalización.
- Esta misma herramienta ha permitido ver qué puntos de facturación sufren una mayor penalización por culpa del $\cos \phi$. Se está procediendo al dimensionamiento de los condensadores para la corrección de este factor y la reducción del coste de la factura.

7.3. Líneas futuras

El resultado de la investigación de esta tesis abre un sinfín de líneas de trabajo. A continuación se tratará de enumerar algunas de las que se consideran más importantes:

- Plantear nuevas herramientas en *Processing* que funcionen en tiempo real. Para ello, las herramientas deberán ser capaces de leer en tiempo real los últimos datos de la base y realizar la proyección del nuevo dato. De todas las técnicas de proyección analizadas, únicamente el CCA sería capaz de proyectar nuevos puntos sin necesidad de volver a entrenar el resto de la proyección. Cabe destacar que los algoritmos de proyección son demasiado costosos computacionalmente como para entrenar todos los puntos en cada iteración. Una posible solución podría ser utilizar aproximaciones que son capaces de proyectar nuevos puntos sin necesidad de entrenar todos, aunque la precisión se reduzca. Por tanto, un trabajo futuro sería validar la precisión de estos métodos de proyección de nuevos puntos sin necesidad de reentrenar y utilizarlos para desarrollar herramientas interactivas en tiempo real.
- Mejorar los aspectos relativos a la visualización en las herramientas de *Processing*. Algunas de las herramientas desarrolladas presentan códigos de colores idénticos para codificar variables que no están relacionadas. El estudio de nuevas escalas de color que permitan seguir diferenciando los distintos valores mostrados y no utilicen colores de otras escalas también utilizadas. Así mismo también buscar formas de mejorar la interactividad del sistema para que sea más sencillo la extracción del conocimiento.
- Volver a generar los modelos del SOM con un mayor número de datos, al menos 5 años, con el fin de aumentar la precisión de los modelos. En esta tesis se han utilizado datos de solo 13 meses por lo que la precisión de los modelos aumentaría al utilizar un mayor número de datos de entrada, es decir, generar modelos que contemplen un mayor número de situaciones posibles.
- Probar otras técnicas de predicción que no estén basadas en técnicas de reducción de la dimensionalidad. En esta tesis se utilizaron estas técnicas por acotar el campo de estudio y aprovechar el conocimiento extraído en la comparación de perfiles. Aunque en la bibliografía no hay un acuerdo claro, se podrían probar técnicas basadas en *Multilayer Perceptron* (MLP), en el dominio de la frecuencia, análisis estadísticos, etc. y compararlas con las técnicas de predicción ya obtenidas en esta tesis.
- Medir otras variables ambientales, como puede ser velocidad y dirección del viento, presión atmosférica, etc. con el fin de estudiar su correlación con el consumo, y así conseguir una mayor precisión en el análisis de la dependencia de la curva de consumo diaria con las condiciones ambientales.
- Probar otras técnicas para el estudio de la dependencia de la variable de consumo con la hora del día. La técnica de correlaciones utilizada en esta tesis ha demostrado ser de utilidad para el caso de los edificios de la Universidad de León, pero carece de cierta robustez por lo que puede ser menos eficiente en otras situaciones. Por ello sería necesario estudiar la viabilidad de otro tipo de técnicas.
- Modificar el algoritmo de obtención del **índice de actividad**, estudiando otras variables que puedan dar idea de la ocupación del edificio. Artículos recientes utilizan, por ejemplo, el nivel de uso de la señal inalámbrica, que es un indicador del nivel de

actividad y por tanto se puede correlacionar con el consumo de potencia. Se podría aumentar la precisión del índice de actividad en función del número de clases que se estén impartiendo mediante un análisis detallado de los horarios. También se podría mejorar mucho este índice si el acceso a las facultades se hiciese por medio de tarjetas RFID, con lo que en todo momento se conocería el número de alumnos y profesores en cada edificio y se tendría una medida real de la actividad. Este conocimiento no solo permitiría mejorar la precisión de los modelos, sino que permitiría mejorar la eficiencia del consumo al relacionar consumos con la cantidad de gente que está dentro del edificio y sería posible automatizar la adquisición de estas variables para generar un índice de actividad dinámico.

- Completar un plan de actuación que aproveche el conocimiento extraído para optimizar el consumo eléctrico. Las líneas maestras de este plan se deben dirigir hacia la reorganización del sistema de suministro de electricidad, supresión de puntos de facturación o conexión a la red eléctrica, eliminación de transformadores con pequeñas cargas, control de los sistemas de refrigeración, aire acondicionado e iluminación de zonas de tránsito comunes y elaboración de informes anuales para la negociación con las compañías eléctricas.
- Automatizar el sistema de predicción de tal manera que los equipos con grandes consumos puntuales (arranque de climatizadores, centrifugadores, hornos, ...) tuviesen limitado su uso en los momentos predichos de alto consumo. De esta forma, de acuerdo a las predicciones dadas por las herramientas, cuando existiese riesgo de sobrepasar la potencia máxima contratada, estos equipos no podrían ponerse en funcionamiento, quedando su arranque programado para más adelante, cuando el consumo del edificio fuera menor.
- Interconectar todos los transformadores de los edificios en anillo para tener en marcha solamente el número de transformadores que necesita el campus y evitar así que los transformadores trabajen al 50 % de su punto óptimo. Puesto que hay edificios con transformadores sobredimensionados y la mayoría solo demandan el máximo de potencia durante unas horas al día, la mayoría de los transformadores están trabajando muy por debajo de su punto óptimo. Una interconexión permitiría tener funcionando un número determinado de transformadores en cada instante de tiempo en su punto óptimo. Sería interesante el estudio de un anillo de interconexión de todos los transformadores del campus con todos los edificios, de tal forma que solo entrarían los transformadores necesarios en función de la predicción realizada por las herramientas planteadas en esta tesis.
- Aprovechar toda esta infraestructura creada, junto con las herramientas como base para el análisis, comparación y predicción de perfiles de consumo, con otro tipo de variables, como puede ser el consumo de gas en sistemas de calefacción. Solo sería necesario cambiar la variable a procesar, puesto que la infraestructura de adquisición ya está preparada para leer datos de calderas. Para ello bastaría indicar a las herramientas la nueva variable y realizar un análisis similar al realizado con la variable eléctrica.

Bibliografía

- Abraham, A. y B. Nath. “A neuro-fuzzy approach for modelling electricity demand in victoria”. *Applied Soft Computing*, 1(2), 127 – 138, 2001.
- Ahola, J., E. Alhoniemi y O. Simula, “Monitoring industrial processes using the self-organizing map”. En *SMCia/99 Proceedings of the 1999 IEEE Midnight-Sun Workshop on Soft Computing Methods in Industrial Applications*, 22–27, IEEE Service Center, Piscataway, NJ, 1999.
- Alonso, S., *Supervisión de la energía eléctrica en edificios públicos de uso docente basada en técnicas de minería de datos visual*. Tesis Doctoral. Universidad de Oviedo, Oviedo, España, 2012.
- Alpaydin, E. *Introduction to Machine Learning*. MIT Press, Cambridge, MA, 2004.
- Amjady, N. “Short-term bus load forecasting of power systems by a new hybrid method”. *IEEE Transactions on Power Systems*, 22(1), 333 –341, feb 2007.
- Armstrong, J. y F. Collopy. “Error measures for generalizing about forecasting methods: Empirical comparisons”. *International Journal of Forecasting*, 8(1), 69 – 80, 1992.
- Asbury, C. “Weather load model for electric demand and energy forecasting”. *IEEE Transactions on Power Apparatus and Systems*, 94(4), 1111 – 1116, jul 1975.
- Barreto, G. A., “Time series prediction with the Self-Organizing Map: A Review.” En P. Hitzler y B. Hammer (editores), *Perspectives on Neural-Symbolic Integration*, Springer-Verlag, 2007.
- Barreto, G. A., J. C. M. Mota, L. G. M. Souza y R. A. Frota, “Nonstationary time series prediction using local models based on competitive neural networks”. En *IEA/AIE 2004: Proceedings of the 17th international conference on Innovations in applied artificial intelligence*, 1146–1155, Springer Verlag Inc., 2004. ISBN 3-540-22007-0.
- Belkin, M. y P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering”. En *Advances in Neural Information Processing Systems*, tomo 14, 585–591, MIT Press, 2001.
- Bellman, R. *Adaptive control processes: a guided tour*. Princeton University Press, 1961.

- Bertoldi, P. y B. Atanasiu, “Electricity consumption and efficiency trends in European Union”. Informe técnico, European Commission. Joint Research Centre. Institute for Energy, 2009.
- Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press, Great Clarendon St, Oxford OX2 6DP, UK., 1995.
- Bishop, C. M. *Pattern recognition and machine learning*, 1ª edición. Information science and statistics. Springer New York, 2006. ISBN 978-0-387-31073-2.
- Bontempi, G., “Long term time series prediction with multi-input multi-output local learning”. En *Proceedings of the 2nd European Symposium on Time Series Prediction (TSP)*, 145–154, 2008.
- Box, G. y G. Jenkins. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.
- Bushby, S. T. “BACnet: a standard communication infrastructure for intelligent buildings”. *Automation in Construction*, 6(6), 529 – 540, 1997.
- Cao, L. “Support vector machines experts for time series forecasting”. *Neurocomputing*, 51(0), 321 – 339, 2003.
- Carpinteiro, O. A. S., A. P. A. da Silva y C. H. L. Feichas, “Hierarchical neural model in short-term load forecasting”. En *Proceedings of the International Joint Conference on Neural Networks*, tomo 6, 241–246, IEEE, Piscataway, NJ, 2000.
- Carreira-Perpiñán, M. Á., *Continuous Latent Variable Models for Dimensionality Reduction and Sequential Data Reconstruction*. Tesis Doctoral. Department of Computer Science, University of Sheffield, UK, 2001.
- Carreira-Perpiñán, M. Á., “A review of dimension reduction techniques”. Informe Técnico CS-96-09, Dept. of Computer Science, University of Sheffield, Diciembre 1996.
- Chappell, G. J. y J. G. Taylor. “The Temporal Kohonen Map”. *Neural Networks*, 6, 441–445, 1993.
- Chatfield, C. y M. Yar. “Holt-Winters forecasting: Some practical issues”. *The Statistician*, 37(2), 129 –140, 1988.
- Chen, J., S.-J. Deng y X. Huo. “Electricity price curve modeling and forecasting by manifold learning”. *IEEE Transactions on Power Systems*, 23(3), 877 –888, 2008.
- Cherkassky, V. y F. Mulier. *Learning from Data – Concepts, Theory and Methods*, 2ª edición. John Wiley & Sons, New York, 2007.
- Cho, J., J. C. Príncipe, D. Erdogmus y M. A. Motter. “Modeling and inverse controller design for an unmanned aerial vehicle based on the self-organizing map.” *IEEE Trans. Neural Networks*, 17(2), 445–460, 2006.

- Contreras, J., R. Espinola, F. Nogales y A. Conejo. "ARIMA models to predict next-day electricity prices". *Power Systems, IEEE Transactions on*, 18(3), 1014 – 1020, aug 2003.
- Cook, J., I. Sutskever, A. Mnih y G. Hinton, "Visualizing similarity data with a mixture of maps." En *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics.*, tomo 2, 67–74, 2007.
- Coombs, K. "Drupal done right". *Library Journal*, 134(19), 2009. Cited By (since 1996): 2.
- Cortes, C. y V. Vapnik. "Support-vector networks". *Machine Learning*, 20(3), 273–297, 1995.
- Dahhaghchi, I., R. Christie, G. Rosenwald y C.-C. Liu. "AI application areas in power systems". *IEEE Expert*, 12(1), 58 –66, jan/feb 1997.
- Davies, D. L. y D. W. Bouldin. "A cluster separation measure". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224 – 227, apr 1979.
- Demartines, P. y J. Héroult. "Curvilinear component analysis: a self organizing neural network for non linear mapping of data sets". *IEEE Transactions on Neural Networks*, 8, 148–154, 1997.
- Díaz, I., *Detección e identificación de fallos en procesos industriales mediante técnicas de procesamiento digital de señal y redes neuronales. Aplicación al mantenimiento predictivo de accionamientos eléctricos*. Tesis Doctoral. Universidad de Oviedo, Oviedo, España, 2000.
- Díaz, I., A. B. Díez y A. A. Cuadrado, "Complex process visualization through continuous feature maps using radial basis functions." En *Proceedings of ICANN 2001, International Conference on Artificial Neural Networks*, tomo 2130, 443–449, Vienna, Austria, 2001.
- Díaz, I. y J. Hollmén, "Residual generation and visualization for understanding novel process conditions." En *Proceedings of International Joint Conference on Neural Networks (IJCNN 2002)*, tomo 3, 2070–2075, IEEE, Piscataway, 2002.
- Dijkstra, E. W. "A note on two problems in connexion with graphs". *Numerische Mathematik*, 1, 269–271, 1959.
- Domínguez, M., *Supervisión remota de procesos complejos vía Internet mediante Técnicas de Data Mining Visual. Aplicación a una planta piloto industrial*. Tesis Doctoral. Universidad de Oviedo, Oviedo, España, 2003.
- Domínguez, M., J. Fuertes, I. Díaz, A. Cuadrado, S. Alonso y A. Morán, "Analysis of electric power consumption using self-organizing maps." En *Proceedings of the 18th World Congress of the International Federation of Automatic Control, IFAC2011*, Milán, 2011.
- Domínguez, M., J. J. Fuertes, S. Alonso, M. A. Prada, A. Morán y P. Barrientos, "Advanced tools for electricity consumption monitoring". En *International Conference on Renewable Energies and Power Quality (ICREPQ12)*, 2012a.

- Domínguez, M., J. J. Fuertes, I. Díaz, M. A. Prada, S. Alonso y A. Morán. “Monitoring industrial processes with SOM-based dissimilarity maps”. *Expert Systems with Applications*, 39(8), 7110 – 7120, 2012b.
- Domínguez, M., P. Reguera y J. J. Fuertes. “Laboratorio Remoto para la enseñanza de la Automática en la Universidad de León (España)”. *Revista Iberoamericana de Automática e Informática Industrial*, 2(2), 36–45, 2005.
- Domínguez, M., P. Reguera, J. J. Fuertes, M. A. Prada, S. Alonso, A. Morán y D. Fernández, “Development of an educational tool in labview and its integration in remote laboratory of automatic control.” En *Advances in Control Education*, tomo 8, Kumamoto, 2009.
- Dubrawsky, I., C. T. Baumrucker, J. Caesar, M. Krishnamurthy, T. W. Shinder, B. Pinkard, E. Seagren *et al.* *Designing and Building Enterprise DMZs*. Elsevier Inc., 2006.
- Dunn, J. C. “Well-separated clusters and optimal fuzzy partitions”. *Journal of Cybernetics*, 4(1), 95–104, 1974.
- Eckerson, W. W. “Three tier client/server architectures: Achieving scalability, performance, and efficiency in client/server applications”. *Open Information Systems*, 3(20), 46–50, 1995.
- Erwin, E., K. Obermayer y K. Schulten. “Self-organizing maps: Order, convergence properties and energy functions”. *Biol. Cyb.*, 67, 47–55, 1992.
- España, *LEY 54/1997, de 27 de Noviembre, del Sector Eléctrico.*, 1997.
- Fan, S. y L. Chen. “Short-term load forecasting based on an adaptive hybrid method”. *IEEE Transactions on Power Systems*, 21(1), 392 – 401, feb. 2006.
- Fan, S., L. Chen y W.-J. Lee. “Short-term load forecasting using comprehensive combination based on multimeteorological information”. *IEEE Transactions on Industry Applications*, 45(4), 1460–1466, 2009.
- Fan, S., C. Mao y L. Chen. “Electricity peak load forecasting with self-organizing map and support vector regression”. *IEEE Transactions on Electrical and Electronic Engineering*, 1(3), 330–336, 2006.
- Fay, D., J. V. Ringwood, M. Condon y M. Kelly. “24-h electrical load data-a sequential or partitioned time series”. *Neurocomputing*, 55(3 4), 469 – 498, 2003.
- Fayyad, U. M., G. Piatetsky-Shapiro y P. Smyth. “From data mining to knowledge discovery in databases”. *AI Magazine*, 17(3), 37–54, 1996.
- Ferreira de Oliveira, M. C. y H. Levkowitz. “From visual data exploration to visual data mining: A survey”. *IEEE Transactions on Visualization and Computer Graphics*, 9(3), 378–394, 2003.

- Figueiredo, V., F. Rodrigues, Z. Vale y J. Gouveia. “An electric energy consumer characterization framework based on data mining techniques”. *IEEE Transactions on power systems*, 20(2), 596–602, 2005.
- Fray, B. J., *Computational information design*. Tesis de Master. Massachusetts Institute of Technology, Cambridge, Massachusetts, 2004.
- Freeman, J. y D. Passmore, “The virtual LAN technology”. Informe Técnico 200374-001, Decisys, Inc., Sterling, VA, 1996.
- Fritzke, B., “Unsupervised ontogenetic networks”. En E. Fiesler y R. Beale (editores), *Handbook of Neural Computation*, Institute of Physics Publishing and Oxford University Press, 1996.
- Fuertes, J. J., *Modelado y representación de la dinámica de sistemas complejos mediante técnicas de Data Mining visual para la supervisión remota de procesos industriales vía Internet*. Tesis Doctoral. Universidad de Oviedo, Oviedo, España, 2006.
- Fuertes, J. J., M. Domínguez, S. Alonso, M. A. Prada, A. Morán y P. Barrientos, “Sistema de supervisión de energía eléctrica en la Universidad de León”. En *Proceedings of the XXXII Jornadas de Automática, JA2011*, 2011.
- Fuertes, J. J., M. A. Prada, M. Domínguez, P. Reguera, I. Díaz y A. A. Cuadrado, “Modeling of dynamics using process state projection on the Self Organizing Map.” En *International Conference on Artificial Neural Networks 2007, Part I, Lecture Notes in Computer Science*, tomo 4668, 589–598, Springer-Verlag, 2007. ISBN 978-3-540-74689-8.
- Fuertes, J. J., P. Reguera, M. Domínguez, I. Díaz y A. A. Cuadrado, “Industrial supervision system based on visual Data Mining and motion trajectory analysis.” En *International Federation of Automatic Control 16th IFAC World Congress*, Prague, 2005.
- García, J., I. Díaz, I. Álvarez, D. Pérez, D. Ordoñez y M. Domínguez, “Spectrogram analysis using manifold learning”. En *Proceedings of the 5th International Scientific Conference on Physics and Control, PHYSCON2011*, 2011.
- Garrett, J. *et al.*, “Ajax: A new approach to web applications”. Informe técnico, Adaptive path, 2005.
- Goh, C. y R. Law. “Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention”. *Tourism Management*, 23(5), 499 – 510, 2002.
- González, P. y J. Zamarreño. “Prediction of hourly energy consumption in buildings based on a feedback artificial neural network”. *Energy and Buildings*, 37, 595–601, 2005.
- Gooijer, J. G. D. y R. J. Hyndman. “25 years of time series forecasting”. *International Journal of Forecasting*, 22(3), 443 – 473, 2006.
- Gordon, A. “A review of hierarchical classification”. *Journal of the Royal Statistical Society (A)*, 150(2), 119–137, 1987.

- Gray, R. M. y D. L. Neuhoff. “Quantization”. *IEEE Transactions on Information Theory*, 44(6), 1–63, 1998.
- Gupta, P. y K. Yamada. “Adaptive short-term forecasting of hourly loads using weather information”. *IEEE Transactions on Power Apparatus and Systems*, PAS-91(5), 2085 – 2094, sept. 1972.
- Guyon, I. y A. Elisseeff. “An introduction to variable and feature selection”. *J. Mach. Learn. Res.*, 3, 1157–1182, Marzo 2003.
- Hagan, M. T. y S. M. Behr. “The time series approach to short term load forecasting”. *IEEE Transactions on Power Systems*, 2(3), 785 –791, Agosto 1987.
- Halkidi, M., Y. Batistakis y M. Vazirgiannis. “On clustering validation techniques”. *Intelligent Information Systems*, 17(2), 107–145, 2001.
- Hammer, B., A. Micheli, A. Sperduti y M. Strickert. “Recursive self-organizing network models.” *Neural Networks*, 17, 1061–1085, 2004.
- Hand, D. J., P. Smyth y H. Mannila. *Principles of data mining*. MIT Press Cambridge, 2001. ISBN 0-262-08290-X.
- Harjono, J., G. Ng, D. Kong y J. Lo, “Building smarter web applications with HTML5”. En *Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research, CASCON'10*, 402–403, 2010.
- Haykin, S. *Neural Networks: A Comprehensive Foundation*. Macmillan Co., New York, 1994. ISBN 0-13-273350-1.
- Hérault, J., C. Jausions-Picaud y A. Guérin-Dugué, “Curvilinear Component Analysis for high-dimensional data representation: I. Theoretical aspects and practical use in the presence of noise”. En *Engineering Applications of Bio-Inspired Artificial Neural Networks*, tomo 1607 de *Lecture Notes in Computer Science*, 625–634, Springer Berlin / Heidelberg, 1999. ISBN 978-3-540-66068-2, 10.1007/BFb0100530.
- Hinneburg, A., D. A. Keim y M. Wawryniuk. “Using projections to visually cluster high-dimensional data”. *IEEE Computing in Science and Engineering*, 5(2), 14–25, 2003.
- Hinton, G. E. y S. Roweis, “Stochastic Neighbor Embedding.” En *Advances in Neural Information Processing Systems*, tomo 15, 833–840, MIT Press, Cambridge, MA, 2002.
- Hippert, H., C. Pedreira y R. Souza. “Neural networks for short-term load forecasting: a review and evaluation”. *IEEE Transactions on Power Systems*, 16(1), 44 –55, feb 2001.
- Hoffman, D., D. Prabhakar y P. Strooper, “Testing IPtables”. En *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research, CASCON '03*, 80–91, IBM Press, 2003.
- Hoga, C. y G. Wong, “IEC 61850: open communication in practice in substations”. En *Power Systems Conference and Exposition, 2004. IEEE PES*, 618 – 623 vol.2, 2004.

- Hollmén, J., *Process modeling using the Self-Organizing Map*. Tesis de Master. Helsinki University of Technology, Espoo, Finland, 1996.
- Hollmén, J. y O. Simula, "Prediction models and sensitivity analysis of industrial production process parameters by using the Self-Organizing Map." En *Proceedings of IEEE Nordic Signal Processing Symposium(NORSIG96)*, 79–82, 1996.
- Hotelling, H. "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 15, 417–441, Sep 1933.
- Hyndman, R. J. y A. B. Koehler. "Another look at measures of forecast accuracy". *International Journal of Forecasting*, 22(4), 679 – 688, 2006.
- Instituto para la Diversificación y Ahorro de la Energía, I., *Plan de Acción 2008-2012 de la Estrategia de Ahorro y Eficiencia Energética en España.*, 2007.
- Ismail, S., A. Shabri y R. Samsudin. "A hybrid model of Self-Organizing Maps (SOM) and Least Square Support Vector Machine (LSSVM) for time-series forecasting". *Expert Systems with Applications*, 38, 10574–10578, August 2011.
- Jacobs, R. A., M. I. Jordan, S. J. Nowlan y G. E. Hinton. "Adaptive mixtures of local experts". *Neural Computation*, 3(1), 79–87, Febrero 1991.
- Johnson, D. H. y S. Sinanovic. "Symmetrizing the kullback-leibler distance". *IEEE Transactions on Information Theory*, 2001.
- Kalogirou, S. A. "Applications of artificial neural-networks for energy systems". *Applied Energy*, 67(1-2), 17 – 35, 2000.
- Kaski, S., *Data Exploration Using Self-Organizing Maps*. Tesis Doctoral. Helsinki University of Technology, Espoo, Finland, 1997.
- Kaski, S., J. Nikkila, M. Oja, J. Venna, P. Toronen y E. Castren. "Trustworthiness and metrics in visualizing similarity of gene expression". *BMC Bioinformatics*, 4(1), 48, 2003.
- Kasslin, M., J. Kangas y O. Simula, "Process state monitoring using Self-Organizing Maps". En I. Aleksander y J. Taylor (editores), *Artificial Neural Networks*, 2, tomo II, 1531–1534, North-Holland, Amsterdam, Netherlands, 1992.
- Keim, D. A. "Visual exploration of large data sets." *Communications ACM*, 44(8), 38–44, 2001.
- Keim, D. A. "Information visualization and visual data mining." *IEEE Trans. Vis. Comput. Graph.*, 8(1), 1–8, 2002.
- Keim, D. A., J. Kohlhammer, G. Ellis y F. Mansmann. *Mastering the information age-solving problems with visual analytics*. Florian Mansmann, 2010.
- Keim, K. H.-P., D. "Visualization techniques for mining large databases: a comparison." *IEEE transactions on knowledge and data engineering*, 8(6), 923–938, 1996.

- Khan, A. "Monitoring power for the future". *Power Engineering Journal*, 15(2), 81 – 85, apr 2001.
- Kim, D. y C. Kim. "Forecasting time series with genetic fuzzy predictor ensemble". *IEEE Transactions on Fuzzy Systems*, 5(4), 523 –535, nov 1997.
- Kline, M. *Methods for multi-step time series forecasting with neural networks*. Information Science Publishing, 2004.
- Kohavi, R., "A study of cross-validation and bootstrap for accuracy estimation and model selection". En *Proceedings of the 14th International joint Conference on artificial intelligence (IJCAI)*, 1995, tomo 2, 1995.
- Kohonen, T. "Self-organizing formation of topologically correct feature maps". *Biol. Cyb.*, 43(1), 59–69, 1982.
- Kohonen, T., "Improved versions of Learning Vector Quantization". En *Proc. IJCNN-90, International Joint Conference on Neural Networks, San Diego*, tomo I, 545–550, IEEE Service Center, Piscataway, NJ, 1990a.
- Kohonen, T., "The Self-Organizing Map." En *Proceedings of the IEEE*, tomo 78, 1464–1480, 1990b.
- Kohonen, T. *Self-Organizing Maps*, 3ª edición. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001. ISBN 3–5406–7921–9.
- Kohonen, T., E. Oja, O. Simula, A. Visa y J. Kangas. "Engineering applications of the Self-Organizing Map". *Proceedings of the IEEE*, 84(10), 1358–84, 1996.
- Koskela, T., M. Varsta, J. Heikkonen y K. Kaski. "Temporal sequence processing using Recurrent SOM". *Proceedings Second International Conference on Knowledge-Based Intelligent Electronic Systems*, 1, 290–297, Abril 1998.
- Kruskal, J. B. y M. Wish. "Multidimensional scaling". *Sage university paper series on quantitative application in the social sciences*, 7–11, 1978.
- Lee, J. A., A. Lendasse, N. Donckers y M. Verleysen, "A robust nonlinear projection method." 13–20, 2000.
- Lee, J. A., A. Lendasse y M. Verleysen. "Nonlinear projection with curvilinear distances: Isomap versus Curvilinear Distance Analysis". *Neurocomputing*, 57, 49–76, Marzo 2004.
- Lee, J. A. y M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer, 2007. ISBN 978-0387393506.
- Lee, J. M. *Introduction to Topological Manifolds*. Springer, 2000. ISBN 978-0-387-98759-0.
- Lendasse, A., J. Lee, V. Wertz y M. Verleysen. "Forecasting electricity consumption using nonlinear projection and self-organizing maps". *Neurocomputing*, 48(1), 299–311(13), Octubre 2002.

- Leung, H., T. Lo y S. Wang. "Prediction of noisy chaotic time series using an optimal Radial Basis Function neural network". *IEEE Transactions on Neural Networks*, 12(5), 1163–1172, sep 2001.
- Lin, G.-F. y L.-H. Chen. "Time series forecasting by combining the Radial Basis Function network and the Self-Organizing Map". *Hydrological Processes*, 19(10), 1925–1937, 2005.
- Linde, Y., A. Buzo y R. Gray. "An algorithm for vector quantizer design". *IEEE Transactions on Communications*, COM-28, 84–95, 1980.
- Ljung, L. *System Identification: Theory for the User, Second Edition*. Prentice-Hall, Englewood Cliffs, NJ, 1999.
- Lobo, V., *Ship noise classification: a contribution to prototype based classifier design*. Tesis Doctoral. New University of Lisbon, Lisboa, 2002.
- MacQueen, J., "Some methods for classification and analysis of multivariate observations". En L. LeCam y J. Neyman (editores), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I: Statistics*, 281–297, University of California Press, Berkeley and Los Angeles, CA, 1967.
- Mandal, P., T. Senjyu, N. Urasaki y T. Funabashi. "A neural network based several-hour-ahead electric load forecasting using similar days approach". *International Journal of Electrical Power & Energy Systems*, 28(6), 367–373, 2006.
- Marcuse, J., B. Menz y J. Payne. "Servers in SCADA applications". *IEEE Transactions on Industry Applications*, 33(5), 1295–1299, sep/oct 1997.
- Martinetz, T. M., S. G. Berkovich y K. J. Schulten. "Neural-gas network for vector quantization and its application to time-series prediction". *IEEE Transactions on Neural Networks*, 4(4), 558–569, 1993.
- Martínez Álvarez, F., A. Troncoso, J. Riquelme y J. Aguilar Ruiz. "Energy time series forecasting based on pattern sequence similarity". *Knowledge and Data Engineering, IEEE Transactions on*, 23(8), 1230–1243, aug. 2011.
- Metaxiotis, K., A. Kagiannas, D. Askounis y J. Psarras. "Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher". *Energy Conversion and Management*, 44(9), 1525–1534, 2003.
- Milidiú, R. L., R. J. Machado y R. P. Rentería. "Time series forecasting through wavelets transformation and a mixture of expert models". *Neurocomputing*, 28(1), 145–156, 1999.
- Milligan, G. W. y M. C. Cooper. "An examination of procedures for determining the number of clusters in a data set". *Psychometrika*, 50(2), 159–179, 1985.
- Modbus Organization, *Modbus Application Protocol Specification*, 2006.

- Morán, A., S. Alonso, M. A. Prada, R. García y J. J. Fuertes, “Time series for fault detection in an industrial pilot plant”. En *Proceedings of the 5th International Scientific Conference on Physics and Control, PHYSCON11*, 2011.
- Morán, A., J. J. Fuertes, A. A. Cuadrado, M. A. Prada, S. Alonso y M. Domínguez, “Análisis de perfiles de consumo energético en edificios de uso docente utilizando técnicas de reducción de la dimensión.” En *Proceedings of the VIII Simposio CEA de Control Inteligente*, 2012a.
- Morán, A., J. J. Fuertes, M. Domínguez, M. A. Prada, S. Alonso y P. Barrientos, “Visual continuous map for electrical bill comparisons”. En *Proceedings of the 10th Portuguese Conference on Automatic Control*, 2012b.
- Morán, A., J. J. Fuertes, M. A. Prada, S. Alonso, P. Barrientos, M. Domínguez y I. Díaz, “Analysis of electricity consumption profiles by means of dimensionality reduction techniques”. En C. Jayne, S. Yue y L. Iliadis (editores), *EANN 2012 - CCIS: Volume number 311*, Communications in Computer and Information Science, Springer Berlin / Heidelberg, 2012c.
- Morán, A., M. A. Prada, S. Alonso, P. Barrientos, J. J. Fuertes, M. Domínguez y I. Díaz, “Two-stage approach for electricity consumption forecasting in public buildings”. En J. Hollmén, F. Klawonn y A. Tucker (editores), *Advances in Intelligent Data Analysis XI*, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2012d.
- Moshtaghi, M., T. C. Havens, J. C. Bezdek, L. Park, C. Leckie, S. Rajasegarar, J. M. Keller *et al.* “Clustering ellipses for anomaly detection”. *Pattern Recognition*, 44(1), 55 – 69, 2011.
- Müller, K., A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen y V. Vapnik, “Predicting time series with Support Vector Machines”. En *Artificial Neural Networks ICANN 97*, tomo 1327 de *Lecture Notes in Computer Science*, 999–1004, Springer Berlin / Heidelberg, 1997. ISBN 978-3-540-63631-1, 10.1007/BFb0020283.
- Murray, F. T., J. V. Ringwood y P. C. Austin. “Integration of multi-time-scale models in time series forecasting”. *International Journal of Systems Science*, 31(10), 1249–1260, 2000.
- Nagata, T., “An electric power energy monitoring system in campus using an Internet”. En *IEEE Power Engineering Society General Meeting*, 1–6, 2006.
- Ngan, H. Y., G. K. Pang y N. H. Yung. “Ellipsoidal decision regions for motif-based patterned fabric defect detection”. *Pattern Recognition*, 43(6), 2132 – 2144, 2010.
- Oglesby, R. y S. Herold. *VMware ESX Server: Advanced technical design guide (Advanced technical design guide series)*. The Brian Madden Company, 2005. ISBN 0971151067.
- Pearson, K. “On lines and planes of closest fit to systems of points in space”. *Philosophical Magazine*, 2(6), 559 – 572, 1901.

- Pérez-Lombard, L., J. Ortiz y C. Pout. “A review on buildings energy consumption information”. *Energy and Buildings*, 40(3), 394 – 398, 2008.
- Prada, M. A., *Técnicas de extracción del conocimiento basadas en data mining visual para la supervisión de procesos industriales. Análisis de la dinámica basado en mapas auto-organizados*. Tesis Doctoral. Universidad de León, León, España, 2009.
- Príncipe, J., L. Wang y M. Motter. “Local dynamic modeling with Self-Organizing Maps and applications to nonlinear system identification and control”. *Proc. IEEE*, 86(11), 2240–2258, 1998.
- Príncipe, J. C., A. Rathie y J. M. Kuo. “Prediction of chaotic time series with neural networks and the issue of dynamic modeling”. *International Journal of Bifurcation and Chaos*, 2(4), 989–996, 1992.
- Qiu, B., H. B. Gooi, Y. Liu y E. K. Chan. “Internet-based SCADA display system”. *IEEE Computer Applications in Power*, 15(1), 14 –19, jan 2002.
- R. Boné, M. C., “Multi-step-ahead prediction with neural networks: a review”. tomo 134, 97–106, 2002.
- Rahman, S. y R. Bhatnagar. “An expert system based algorithm for short term load forecast”. *IEEE Transactions on Power Systems*, 3(2), 392 –399, may 1988.
- Reas, C. y B. Fry. *Processing: A Programming Handbook for Visual Designers and Artists*. MIT Press, 2007.
- Rossiter, A. “Make your plant more energy efficient”. *Chemical Engineering Progress, CEP*, 101(12), 31–36, 2005.
- Rousseew, P. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. *Journal of Computational and Applied Mathematics*, 20(0), 53 – 65, 1987.
- Roweis, S. T. y L. K. Saul. “Nonlinear dimensionality reduction by Locally Linear Embedding”. *Science*, 290, 2323–2326, 2000.
- Sammon, Jr., J. W. “A non-linear mapping for data structure analysis”. *IEEE Transactions on Computers*, 18, 401–409, 1969.
- Sideratos, G., I. Vitellas y N. Hatziaargyriou, “A load forecasting hybrid method for an isolated power system”. En *16th International Conference on Intelligent System Application to Power Systems (ISAP)*, 1 –5, 2011.
- Simon, G., A. Lendasse, M. Cottrell y M. Verleysen, “Long-term time series forecasting using Self-Organizing Maps: the Double Vector Quantization Method”. En *ANNPR 2003, Artificial Neural Networks in Pattern Recognition, Florence (Italy)*, 8–14, 2003.
- Sorjamaa, A., J. Hao, N. Reyhani, Y. Ji y A. Lendasse. “Methodology for long-term prediction of time series”. *Neurocomputing*, 70(16-18), 2861–2869, Octubre 2007.

- Sorjamaa, A. y A. Lendasse, “Time series prediction using DirRec strategy”. En *ESANN 2006, European Symposium on Artificial Neural Networks*, 143–148, 2006.
- Specht, D. F. “A General Regression Neural Network”. *IEEE Transactions on Neural Networks*, 2(6), 568–576, 1991.
- Suykens, J., J. D. Brabanter, L. Lukas y J. Vandewalle. “Weighted Least Squares Support Vector Machines: robustness and sparse approximation”. *Neurocomputing*, 48(1 - 4), 85 – 105, 2002a.
- Suykens, J., T. van Gestel, J. D. Brabanter, B. D. Moor y J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002b.
- Taieb, S. B., G. Bontempi, A. Atiya y A. Sorjamaa. “A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition”. *Expert Systems with Applications*, 2011.
- Taieb, S. B., A. Sorjamaa y G. Bontempi. “Multiple-output modeling for multi-step-ahead time series forecasting”. *Neurocomputing*, 73(10-12), 1950 – 1957, 2010.
- Tay, F. E. H. y L. J. Cao. “Improved financial time series forecasting by combining support vector machines with self-organizing feature map”. *Intelligent Data Analysis*, 5(4), 339–354, Septiembre 2001.
- Tenenbaum, J. B., V. de Silva y J. C. Langford. “A global geometric framework for nonlinear dimensionality reduction”. *Science*, 290, 2319–2323, 2000.
- Theil, H. *Applied economic forecasting*. North-Holland, 1966.
- Thompson, M. L. “Selection of variables in multiple regression: Part I. a review and evaluation”. *International Statistical Review*, 46(1), 1–19, 1978.
- Tiao, G. C. y R. S. Tsay. “Some advances in non-linear and adaptive modelling in time-series”. *Journal of Forecasting*, 13(2), 109–131, 1994.
- Tryba, V. y K. Goser, “Self-Organizing Feature Maps for process control in chemistry”. En *Artificial Neural Networks*, 847–852, North-Holland, Amsterdam, Netherlands, 1991.
- Tryba, V., S. Metzen y K. Goser, “Designing basic integrated circuits by Self-Organizing Feature Maps”. En *Neuro-Nîmes '89. Int. Workshop on Neural Networks and their Applications*, 225–235, ARC; SEE, EC2, Nanterre, France, 1989.
- Tufte, E. *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- Tukey, J. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, 1977.
- Ultsch, A. y H. P. Siemon, “Kohonen’s Self Organizing Feature Maps for exploratory data analysis”. En *INNC Paris 90*, 305–308, Universitat Dortmund, 1990.
- Unión Europea, *DIRECTIVA 2010/31/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 19 de Mayo relativa a la eficiencia energía de los edificios.*, 2010.

- US Energy Information Administration, E., *International Energy Outlook 2007 (IEO2007)*, 2007.
- Van der Maaten, L. y G. Hinton. “Visualizing data using t-SNE”. *Journal of Machine Learning Research*, 9, 2579–2605, 2008.
- Van der Maaten, L. J. P., E. O. Postma y H. J. v. D. Herik. “Dimensionality reduction : A comparative review”. *Review Literature And Arts Of The Americas*, 10(February), 1–35, 2007.
- Venna, J., *Dimensionality reduction for visual exploration of similarity structures*. Tesis Doctoral. Helsinki University of Technology, Espoo, Finland, 2007.
- Venna, J. y S. Kaski. “Local multidimensional scaling.” *Neural Networks*, 19(6-7), 889–899, Agosto 2006.
- Venna, J. y S. Kaski. “Comparison of visualization methods for an atlas of gene expression data sets”. *Information Visualization*, 6, 139–154, 2007.
- Verveer, P. y R. Duin. “An evaluation of intrinsic dimensionality estimators”. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 17(1), 81 –86, jan 1995.
- Vesanto, J., “Using the SOM and local models in time-series prediction”. En *Proceedings of Workshop on Self-Organizing Maps (WSOM’97)*, 209–214, 1997.
- Vesanto, J. “SOM-based data visualization methods.” *Intelligent Data Analysis*, 3(2), 111–126, 1999.
- Vesanto, J., *Data Exploration Process Based on the Self-Organizing Map*. Tesis Doctoral. Helsinki University of Technology, Espoo, Finland, 2002.
- Vesanto, J. y E. Alhoniemi. “Clustering of the Self-Organizing Map”. *IEEE Transactions on Neural Networks*, 11(3), 586–600, Mayo 2000.
- Vesanto, J., E. Alhoniemi, J. Himberg, K. Kiviluoto y J. Parviainen. “Self-Organizing Map for Data Mining in MATLAB: The SOM toolbox.” *Simulation News Europe*, (25), 54, 1999.
- Voegtlin, T. “Recursive Self-Organizing Maps”. *Neural Networks*, 15(8), 979–991, 2002.
- Walter, J., H. Ritter y K. Schulten, “Non-linear prediction with Self-Organizing Maps”. En *Proc. IJCNN-90, International Joint Conference on Neural Networks, San Diego*, tomo 1, 589–594, IEEE Service Center, Piscataway, NJ, 1990.
- Walters, B. “VMware virtual platform”. *Linux J.*, 1999, Julio 1999.
- Ware, C. *Visual thinking for design*. Morgan Kaufmann, 2008.
- Weigend, A. S. y N. A. Gershenfeld. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison Wesley, Reading, MA, 1993.

- Wong, P. C. “Visual data mining”. *IEEE Computer Graphics and Applications*, 19(5), 20–21, 1999.
- Xu, R. y D. Wunsch. “Survey of clustering algorithms”. *Neural Networks, IEEE Transactions on*, 16(3), 645–678, Mayo 2005.
- Yang, L., “Sammon’s nonlinear mapping using geodesic distances”. En *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, tomo 2, 303 – 306 Vol.2, 2004.
- Yee, P. y S. Haykin. “A dynamic regularized Radial Basis Function network for nonlinear, nonstationary time series prediction”. *Signal Processing, IEEE Transactions on*, 47(9), 2503 –2521, sep 1999.
- Yin, H. “Data visualisation and manifold mapping using ViSOM”. *Neural Networks*, 15, 1005–1016, 2002.
- Zaknich, A. “Introduction to the modified probabilistic neural network for general signal processing applications”. *IEEE Transactions on Signal Processing*, 46(7), 1980 –1990, jul 1998.
- Zhang, G., B. Patuwo y M. Y. Hu. “A simulation study of artificial neural networks for nonlinear time-series forecasting”. *Computers & Operations Research*, 28(4), 381 – 396, 2001.
- Zhang, L., G. Dai, G. Zhai, Y. Cao y Z. Liu, “Load forecasting via manifold regularization”. En *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, 1 –6, 2009.
- Zhong, Z., C. Xu, B. Billian, L. Zhang, S.-J. Tsai, R. Conners, V. Centeno *et al.* “Power system frequency monitoring network (FNET) implementation”. *IEEE Transactions on Power Systems*, 20(4), 1914 – 1921, nov. 2005.